

AL A082393

(12)

LEVEL II



Research Sponsored by

The Office of Naval Research  
under Contract ONR-N00014-75-C-0609

The Naval Research Laboratory  
under Contract NRL-N00014-78-C-0778

and

The National Science Foundation  
under Grant NSF-ENG-77-20799

DTIC  
ELECTE  
S MAR 28 1980 D  
B

UCLA ENG-8001  
JANUARY 1980

ACCESS-CONTROL SCHEMES FOR REAL-TIME  
AND STORE-AND-FORWARD MULTIPLE-ACCESS  
COMMUNICATION CHANNELS

DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

MING YIU LOUIE

Principal Investigator: I. RUBIN

80 3 05 046

UCLA • SCHOOL OF ENGINEERING AND APPLIED SCIENCE

FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED
ACCESS-CONTROL SCHEMES FOR REAL-TIME AND STORE-AND-FORWARD MULTIPLE-ACCESS COMMUNICATION CHANNELS		Technical Report, Interim
7. AUTHOR(s)		6. PERFORMING ORG. REPORT NUMBER
MING YIU / LOUIE Izhak/Rubin		UCLA-ENG - 8001
9. PERFORMING ORGANIZATION NAME AND ADDRESS		8. CONTRACT OR GRANT NUMBER(s)
Department of System Science University of California Los Angeles, CA 90024		NO0014-75-C-0009 NO0014-78-C-0778
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
see Item 18		ONR- P-4398-Rev-79 NRL- P-4540-Rev-80
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE
12 193		JAN 1980
		13. NUMBER OF PAGES
		161
		15. SECURITY CLASS. (of this report)
		UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)		
<p>Reproduction in whole or in part is permitted for any purpose of the United States Government</p> <p style="text-align: right;"><b>DISTRIBUTION STATEMENT A</b> Approved for public release; Distribution Unlimited</p>		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES (from Item 11):		
a) ONR Statistics and Probability Program, Arlington VA 22217		
b) Naval Research Lab, Code 7558, Washington DC 20375		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
Access-Control Schemes -- Real-Time Transmission -- TDMA Schemes -- Store-and-Forward Transmission -- Hybrid TDMA/Collision-Resolving Schemes -- Reservation Schemes -- Pre-emptive Priority Discipline -- Queueing Theory -- Markov Processes		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		
Demand-Assignment TDMA schemes, a class of hybrid TDMA/Collision-Resolving schemes, governing the sharing of multiple-access communication channels are studied. Sources communicate with each other through a synchronized (slotted), fully-connected communication medium. This communication medium can be a terrestrial radio or line communication channel (inducing low propagation delay) or a satellite channel (inducing high propagation delay).		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102 LF 014 6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

405213

Jm

20-7  
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

→ The Demand-Assignment TDMA schemes studied here are used to allocate channel capacity among sources which require real-time transmission. A non-preemptive Cutoff Priority discipline is employed to offer priority services to important messages. The performance of the schemes is measured in terms of the message blocking (loss) probability and the message delay vs. channel throughput functions. For sources which transmit at multiple rates, a maximum normalized average waiting time is introduced as an overall system performance measure. The latter is used as an objective function in finding the optimal channel frame structure.

The class of store-and-forward hybrid TDMA/Collision-Resolving schemes, ~~developed here~~, are composed of a TDMA component and a Tree Search component. Groups of sources are served on a TDMA basis. Collisions among sources within each group are resolved by following a Tree Search technique. Message arrivals in a sequence of slots are assumed to be i.i.d., governed by an arbitrary distribution. The messages are assumed to contain single packets, except in the pure TDMA case where the message length distribution is arbitrary. Two cases are studied. In one case, the source buffers are assumed to have limited capacities, and in the other case, unlimited capacities. A recursive formulation technique is employed to evaluate the performance of these schemes.

→ Fixed Reservation schemes, operating on a store-and-forward basis, are, <sup>also</sup> investigated as well. A preemptive priority discipline is incorporated as an important feature. Message arrivals in a sequence of frames are assumed to be i.i.d. with an arbitrary distribution. The message length distribution is also assumed to be arbitrary. A technique is developed to yield bounds on the average delay of an arbitrary priority class of messages. In some special cases, these bounds converge and yield an exact result. This is demonstrated by a numerical example.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

ACCESS-CONTROL SCHEMES FOR REAL-TIME AND  
STORE-AND-FORWARD MULTIPLE-ACCESS  
COMMUNICATION CHANNELS

by

Ming Yiu Louie

Principal Investigator: IZHAK RUBIN

Research Sponsored by

The Office of Naval Research  
under Contract ONR-N00014-75-C-0609

The Naval Research Laboratory  
under Contract NRL-N00014-78-C-0778

and

The National Science Foundation  
under Grant NSF-ENG77-20799

Reproduction in whole or in part is permitted for  
any purpose of the United States Government

DTIC  
ELECTE  
S MAR 28 1980 D  
B

Department of System Science  
School of Engineering and Applied Science  
University of California  
Los Angeles, CA 90024

## ACKNOWLEDGMENTS

It is a great pleasure for the author to acknowledge the guidance provided by Professor Izhak Rubin during the course of this dissertation. Professor Rubin's encouragement, patience and numerous suggestions throughout the course of this investigation are deeply appreciated. Grateful acknowledgment is also made to Professors Bryan Ellickson, Jim K. Omura, Charles Stone and Kung Yao for serving on the dissertation committee.

Grateful appreciation is expressed for the support provided by the Office of Naval Research under Contract ONR-N00014-75-C-0609, by the Naval Research Laboratory under Contract NRL-N00014-78-C-0778 and by the National Science Foundation under Grant NSF-ENG77-20799.

Special thanks are extended to my sister Susanna for her fruitful suggestions in the preparation of this dissertation.

Sincere appreciation is made to Kathy Campbell and Lucia M. Rodriguez for their excellent typing of this manuscript.

ACCESSION for		
NTIS	White Section	<input checked="" type="checkbox"/>
DDC	Buff Section	<input type="checkbox"/>
UNANNOUNCED		<input type="checkbox"/>
JUSTIFICATION		
<b>PER LETTER</b>		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL	and/or SPECIAL
<b>A</b>		

## ABSTRACT

Demand-Assignment TDMA schemes, a class of hybrid TDMA/Collision-Resolving schemes and Fixed Reservation schemes, governing the sharing of multiple-access communication channels, are studied. Sources communicate with each other through a synchronized (slotted), fully connected communication medium. This communication medium can be a terrestrial radio or line communication channel (inducing low propagation delay) or a satellite channel (inducing high propagation delay).

The Demand-Assignment TDMA schemes studied here are used to allocate channel capacity among sources which require real-time transmission. A non-preemptive Cutoff Priority discipline is employed to offer priority services to important messages. The performance of the schemes is measured in terms of the message blocking (loss) probability and the message delay vs. channel throughput functions. For sources which transmit at multiple rates, a maximum normalized average waiting time is introduced as an overall system performance measure. The latter is used as an objective function in finding the optimal channel frame structure.

The class of store-and-forward hybrid TDMA/Collision-Resolving schemes, developed here, are composed of a TDMA component and a Tree Search component. Groups of sources are served on a TDMA basis. Collisions among sources within each group are resolved by following a Tree Search technique. Message arrivals in a sequence of slots are assumed to be i.i.d., governed by an arbitrary distribution. The messages are assumed to contain single packets, except in the pure TDMA case where the message length distribution is arbitrary. Two cases are studied. In one case, the source buffers are

assumed to have limited capacities and, in the other case, unlimited capacities. A recursive formulation technique is employed to evaluate the performance of these schemes.

Fixed Reservation schemes, operating on a store-and-forward basis, are investigated as well. A preemptive priority discipline is incorporated as an important feature. Message arrivals in a sequence of frames are assumed to be i.i.d. with an arbitrary distribution. The message length distribution is also assumed to be arbitrary. A technique is developed to yield bounds on the average delay of an arbitrary priority class of messages. In some special cases, these bounds converge and yield an exact result. This is demonstrated by a numerical example.

# TABLE OF CONTENTS

	<u>Page</u>
List of Tables .....	vi
List of Figures .....	vii
ACKNOWLEDGEMENTS .....	ix
VITA .....	x
ABSTRACT .....	xi
CHAPTER I INTRODUCTION .....	1
1.1 Introduction .....	1
1.2 History .....	10
1.3 The System Models .....	14
1.4 Outline of the Dissertation .....	14
CHAPTER II DEMAND-ASSIGNMENT TDMA SCHEMES WITH CUTOFF PRIORITY FOR REAL-TIME SOURCES .....	16
2.1 Introduction .....	16
2.2 The System Models .....	16
2.3 The Performance of Demand-Assignment TDMA Schemes .....	20
2.4 The Performance of Demand-Assignment TDMA Schemes with Cutoff Priority .....	21
CHAPTER III DEMAND-ASSIGNMENT TDMA SCHEMES WITH FIXED BOUNDARIES FOR MULTIPLE-RATE REAL-TIME SOURCES .....	33
3.1 Introduction .....	33
3.2 The System Models .....	33
3.3 The Performance of Demand-Assignment TDMA Schemes .....	36
3.4 Optimal Allocation of Slots .....	38
CHAPTER IV A CLASS OF HYBRID TDMA/COLLISION-RESOLVING SCHEMES .....	51
4.1 Introduction .....	51
4.2 The System Models .....	54
4.3 Two Versions of a Tree Search Scheme .....	60
4.4 Collision-Resolving Procedures .....	74
4.5 Sources with Limited Buffer Capacities .....	83
4.6 Sources with Unlimited Buffer Capacities .....	102



	<u>Page</u>
CHAPTER V FIXED RESERVATION SCHEMES WITH PREEMPTIVE PRIORITIES .....	134
5.1 Introduction .....	134
5.2 The System Models .....	135
5.3 Bounds on the Average Message Delay .....	141
5.4 Fixed Reservation Schemes with Preemptive Priorities .....	156
CHAPTER VI CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH .....	172
6.1 Conclusions .....	172
6.2 Suggestions for Future Research .....	174
APPENDIX THE APPROXIMATION OF A DEMAND-ASSIGNMENT TDMA SYSTEM BY AN M/M/K QUEUEING SYSTEM .....	175
BIBLIOGRAPHY .....	179

# LIST OF TABLES

Table	Page
2.1 Four ways of treating Requests when immediate slot(s) cannot be granted.....	23
4.1 $P_{T_L 2}(k 2)$ under Policy 1.....	84
4.2 $E[W_L 2]$ under Policy 1.....	84
4.3 $P_{T_L 2}(k 2)$ under Policy 2.....	85
4.4 $E[W_L 2]$ under Policy 2.....	85

# LIST OF FIGURES

Figure	Page
2.1 TDMA Time Frame Structure.....	18
2.2 $\ell_1$ vs $\ell_2$ Curves of Case I.....	28
2.3 $P(W_1 > 0)$ vs $\ell_2$ Curves of Case II.....	29
2.4 $\ell_1$ vs $P(W_2 > 0)$ Curves of Case III.....	30
3.1 Slot Allocation under DA/TDMA Schemes with Fixed Boundaries..	47
3.2 $\text{Max}(\tilde{W}_1, \tilde{W}_1)$ vs $\lambda_2$ Curves under DA/TDMA Schemes with Fixed Boundaries.....	50
4.1 Time Frame Structure for a Class of Hybrid Access-Control Schemes.....	57
4.2 The Tree Structure of a Group of Sources under Policy 1, $L = 4$ , with an Example illustrating the Selection of Successive Nodes.....	62
4.3 An Example of Superframe Construction. Propagation Delay is $R$ Slots and $2\alpha < R < 3\alpha$ . ....	65
4.4 The Tree Structure of a Group of Sources under Policy 2, $L = 4$ , with an Example illustrating the Selection of Successive Nodes. ....	69
4.5 An Example illustrating the Effect of Propagation Delay under Policy 1 and Policy 2. $N_G$ and $R$ are of Fixed Values. Under Policy 1, $\alpha = N_G$ and $\alpha < R < 2\alpha$ . Under Policy 2, $\alpha = 2N_G$ and $R + 1 < \alpha$ .....	73
4.6 Collision-Resolving Procedures under Policy 1 when $L > 1$ ....	77
4.7 Collision-Resolving Procedures under Policy 2 when $L > 1$ ....	80
4.8 $E[W^{(2)} + W^{(3)}]$ vs Throughput Curves under Policy 1.....	96
4.9 $\ell$ vs Throughput Curves under Policy 1. ....	97
4.10 Throughput vs $\lambda$ Curves under Policy 1. ....	98
4.11 Delay-Throughput Curves under Policy 1, $R = 0$ . ....	99
4.12 Delay-Throughput Curves under Policy 1, $R = 12$ . ....	100
4.13 $E[W^{(2)} + W^{(3)}]$ vs Throughput Curves under Policy 2. ....	103

4.14	$\lambda$ vs Throughput Curves under Policy 2.....	104
4.15	Throughput vs $\lambda$ Curves under Policy 2.....	105
4.16	Delay-Throughput Curves under Policy 2, $R = 0$ .....	106
4.17	Delay-Throughput Curves under Policy 2, $R = 12$ .....	107
4.18	Delay-Throughput Curves (Lower Envelopes) under Policy 1 and Policy 2, $R = 12$ .....	108
4.19	$E[Q]$ vs Throughput Curves under Policy 1.....	125
4.20	Delay-Throughput Curves under Policy 1, $R = 0$ .....	126
4.21	Delay-Throughput Curves under Policy 1, $R = 12$ .....	127
4.22	$E[Q]$ vs Throughput Curves under Policy 2.....	128
4.23	Delay-Throughput Curves under Policy 2, $R = 0$ .....	129
4.24	Delay-Throughput Curves under Policy 2, $R = 12$ .....	130
4.25	$E[W^{(2)} + W^{(3)}]$ vs Throughput Curves under Policy 2 by Simulation and Approximation, $L = 2$ .....	132
4.26	$E[W^{(2)} + W^{(3)}]$ vs Throughput Curves under Policy 2 by Simulation and Approximation, $L = 3$ .....	133
5.1	An Example illustrating the effect of Propagation Delay on the Broadcast Channel under a Fixed Reservation Scheme. In the First Case, $R_R = R$ and $M < \beta$ . In the Second Case, $R + \gamma \geq R_R \geq R$ and $M^R = \beta$ .....	138
5.2	Class 2 Average Message Delay vs $\rho_2$ under a Preemptive Reservation Scheme.....	170
A	The Frame Latency, $\tau_L$ , and the Holding Time, $T$ , of a Message under a DA/TDMA Scheme, for (a) $W = 0$ and (b) $W > 0$ ..	176

## CHAPTER I

### INTRODUCTION

#### 1.1 Introduction

In the following chapters, multi-accessing of a communication channel by a number of sources is considered. This channel serves as a communication medium between a population of geographically distributed sources. A channel is said to have broadcast capabilities if every source can receive and listen to the messages transmitted by any source, including itself. In order to provide this broadcast capabilities, a single repeater is assumed to be established. Messages arriving at the sources are transmitted in an uplink manner to the repeater. Upon receiving these messages, the repeater then transmits them in a downlink manner (broadcast) to all the sources. Of course, uplink and downlink transmission are carried out in separate frequency bands, otherwise, interference would destroy the messages. A few examples of how a repeater is used in practical situations are described as follows. The repeater can be a satellite transponder in geosynchronous orbit with the earth. The sources sharing the broadcast channel provided by the transponder are the earth stations covered by the satellite (See [1]). In another situation the repeater can be an air-borne (mobile) radio-repeater station providing communication medium between a squadron of military units (sources) (see [2] and [3]). The repeater can also be a business radio-repeater station serving intracity subscribers (sources). All these practical situations serve as the main motivation of our studies. Even though our studies

are motivated by these practical situations, the results of our studies are also applicable to many terrestrial communication networks when considering network channel multi-accessing. It is easily observed that in a radio-repeater situation, the propagation delay of the channel is negligible while in the satellite transponder situation, the round trip propagation delay of the channel is comparatively much longer, about 0.27 seconds. This propagation delay will serve as an important parameter in our studies.

Since the channel is shared among a number of sources, an access-control mechanism is required to appropriately allocate the channel capacity (resource) among the sources in order to make the most use of the capacity of the channel and serve the sources in the most efficient manner. Two methods of carrying out access-control, known as central control and distributed control, are described as follows.

Under a central control scheme, a central controller is established which collects information about the sources, the channel and the messages which are to be transmitted. Based on this collected information, the central controller then appropriately supervises transmission scheduling. This intelligent unit, because of its high complexity, can allow the implementation of very sophisticated and efficient access-control disciplines. However, in order to collect the required information and to send supervision information, not only additional delays are introduced, but also additional protocol messages are generated. Furthermore, the centralization of control increases the vulnerability of the system.

Under a distributed control scheme, management of the channel is shared by all the sources. Therefore, the sources are required to be equipped with more sophisticated instruments so as to carry out their individual control duty and achieve coordination. Such a system is less vulnerable. However, coordination is not easy and more sophisticated equipment at every source induces higher costs.

Three important issues should come to mind when the implementation of an appropriate access-control mechanism is considered. First, the nature, capabilities and geographical distribution of the sources that share the channel. Second, the capacity and the propagation delay of the channel. Third, the characteristics of the messages arriving at the sources which require transmission over the channel.

If messages are arriving at the sources in a deterministic manner and (or) the holding times (occupation times) of the channel for the transmission of messages are deterministic, the task of channel access-control would be much easier. Unfortunately, these factors are usually nondeterministic and can only be described by probability distributions. Furthermore, the flow of message traffic may fluctuate greatly between regular and critical periods which are a priori unknown functions of time, geographic distribution of the sources and situations. Hence, an efficient access-control mechanism is required not only to provide adequate general services to the sources at regular times but is also required to be able to cope with these message traffic fluctuations and offer acceptable services at critical times. Due to the stochastic characteristics of the message traffic, stochastic processes and queueing models are important tools

employed in our studies to evaluate the performance of access-control schemes.

If the message traffic arriving at each source is steady, Fixed-Assignment FDMA (Frequency Division Multiple Access) or Fixed-Assignment TDMA (Time Division Multiple Access) schemes are very efficient (see [4]). Under a Fixed-Assignment FDMA scheme, the bandwidth of the channel is divided into appropriate portions and each source is assigned one or more portions so that each source can transmit its messages in its own assigned portions of the channel bandwidth. Under a Fixed-Assignment TDMA scheme, the channel time is divided into consecutive equal durations called frames. In a fixed manner, each frame is divided into appropriate portions, slots, and each source is assigned one or more slots in each frame so that each source can transmit its messages at full capacity of the channel in its own assigned slots. Under both schemes, each source utilizes only its own assigned portion of the channel, thus, sees the channel as transparent and therefore no contention is developed among the sources. Furthermore, the sizes of the portions can be updated periodically in order to adapt to message traffic fluctuations.

When the message traffic arriving at the sources is bursty (statistically fluctuating), then both Fixed-Assignment FDMA and Fixed-Assignment TDMA schemes yield low utilization of the channel. Consequently, we consider Demand-Assignment FDMA (DA/FDMA) or Demand-Assignment TDMA (DA/TDMA) schemes. If all the sources require transmission at a specific source information rate, that is, each source requires real-time transmission when a subchannel is granted,



we do the following. Under a Demand-Assignment FDMA scheme, the bandwidth of the channel is divided in such a way that each frequency subband forms a subchannel which is adequate to support the message transmission of one single source. Similarly, under a Demand-Assignment TDMA scheme, every channel time frame is divided into equal durations called slots. Each sequence of slots (of fixed frame position) in successive frames forms a subchannel which is adequate to support the message transmission of one single source. For example, the periodic sequence containing the first slot in each frame forms a subchannel. Coded voice transmission applications form a typical example. Usually, under a Demand-Assignment FDMA or Demand-Assignment TDMA scheme, the number of subchannels is less than the total number of sources which share the channel, therefore, subchannels are granted to sources on a demand-assignment basis. Sources which have messages to transmit request for subchannels. To meet these requests, a control mechanism is established to appropriately assign available subchannels to sources. The control can be centralized or distributed. Protocol information for requesting and relinquishing subchannels can be sent through an assigned portion of the channel or a separate channel.

In practice, some messages are more important than others, therefore, we need to study priority schemes. In this dissertation, we employ a Cutoff Priority discipline to give advantage to important messages.

In realistic situations, sources usually do not transmit at one single source information rate. Therefore, we also develop and study

demand-assignment schemes to cover the case in which the sources require real-time transmission at multiple information rates. In order to satisfy their information rates, appropriate allocation of the capacity of the channel with respect to different information rates has to be considered and incorporated into the control mechanism (see [5]).

In many situations, messages are not required to be transmitted at a specific source information rate on a real-time basis. In order to reduce transmission delay and increase throughput, it is more efficient to allow only a few sources to transmit simultaneously at the higher channel transmission rate than to allow many sources to transmit simultaneously at the lower source transmission rate (see [6]). Data transmission applications form a typical example. Such channel sharing schemes are said to operate on a store-and-forward (message-switching or packet-switching) basis. An example of such a store-and-forward procedure is a Polling scheme. Under such a scheme, in a cyclic manner, the sources are scheduled to transmit at full channel capacity. The policy governing the holding time of the channel by each source may vary. However, due to the walk time required in switching from one source to another, Polling may not be efficient. For example, if the number of sources is large and every one generates information in low duty cycle fashion, that is, every one generates short messages interrupted by long pause intervals, a Polling scheme yields poor channel utilization and long message delays.

In many situations, when a packet-switching access-control scheme is used, messages are decomposed into blocks of equal lengths called packets and are transmitted over the channel separately.

Therefore, each packet carries identification and destination information so that the packets can be reassembled at the destination to recover the message. Such packet-switching access-control schemes include various versions of Random Access and reservation procedures.

Under a Random Access scheme (see [7]), a newly generated packet is transmitted without waiting. If two or more packets are transmitted at the same time, a collision occurs. The sources are assumed to have the ability to distinguish if a received broadcast is a successful transmission or a collision. Upon sensing a collision, the collided packets are retransmitted at random future times. Retransmission is repeated until the collided packets get through. Unfortunately, the maximum throughput of the channel implemented by this scheme is very low, equal to  $1/2e$ . To improve the throughput, the Random Access procedure is modified to form a Slotted Random Access scheme (see [8]) by dividing the channel time into slots such that each slot can accommodate the transmission of exactly one packet. Transmission or retransmission of packets are allowed to start only at these slots. Such a scheme requires synchronization of slot times. The maximum throughput is improved to  $1/e$ , but still too low to be acceptable in many practical situations. Furthermore, under both Random Access and Slotted Random Access procedures, when a traffic burst causes most of the transmitted packets to be generated due to retransmission, the channel will become unstable, yielding excessive packet delays and diminishing throughput values.

Under a Tree Search random access scheme (see [9]), the channel time is again slotted as under a Slotted Random Access scheme. New

packets start their transmission at the beginning of specially dynamically recognized time epochs. Each epoch is made up of subperiods called steps. Each step is made up of two slots. At each step, transmission rights are assigned to certain sources. The policy governing the selection of sources to be given transmission rights is as follows. At the first step of an epoch, the entire population is divided into two groups of an equal number of sources. Transmission rights in the two slots are given to the two groups separately. If no collision occurs in either slot, the epoch ends and a new one starts subsequently. Otherwise, transmission rights in successive steps of the epoch are assigned according to the following two rules. First, if a collision occurs in one slot, the group of sources which are given transmission rights in that slot are divided into two smaller groups of an equal number of sources which are then separately given transmission rights in the two slots of the next step. Second, if collisions occur in both slots, the collision in one group is completely resolved before the other is attempted. Both rules are applied to collisions not only at the first step but also at subsequent steps, if necessary, to resolve collision(s). The epoch ends when all the collision(s) and subsequent collision(s), if any, are resolved. A new epoch starts thereafter. Packets which arrive after the start of an epoch are not allowed to be transmitted during the on-going epoch but will be transmitted during the next epoch. Such a scheme is stable, but the maximum throughput is very low, equal to only 0.432.

We observe that under Fixed-Assignment TDMA schemes, in order

to avoid conflicts among sources, every source is assigned one single slot in every channel time frame. Under a random access scheme, actions are taken only when collisions occur. It would be advantageous to integrate the features of these classes of schemes. One such method involves the use of a Group Random Access scheme. Under this scheme, the entire population is divided into groups and the channel time frame is also divided into the same number of portions such that each group of sources is associated with one portion of the channel time frame. The sources in each group are allowed to use only the group's assigned time portion of the channel. Hence, there is no conflict between groups. Conflicts among sources in each group are resolved by the Random Access technique. This scheme has been studied extensively by Rubin [10]. The schemes we have developed and studied work as follows. The entire population is divided into groups containing an equal number of sources. The channel time frame is also divided into the same number of equal portions. Each group is associated with one portion of the channel time frame and the sources of each group are allowed to utilize only the group's associated portion of the channel so that there is no contention between groups. Sources share their group time portions on a random access basis. Conflicts in each group are resolved by the Tree Search technique. By varying the number of groups to be divided from the entire population (consequently varying the number of sources in each group), we obtain a family of hybrid access-control schemes, with the pure Fixed-Assignment TDMA and the pure Tree Search schemes as extreme members of the family.

Under these schemes, a message to be transmitted over the channel is first decomposed into packets; then these packets are transmitted independently over the channel. These schemes exhibit favorable delay-throughput functions if the messages arriving at the sources contain single packets. For long messages which are composed of multiple packets, these schemes will not be efficient because of two reasons. First, the total delay of a message is equal to the sum of the delays of its packets which are transmitted individually and independently over the channel can become excessively long. Second, packet reassembly is necessary at the destination.

To accomodate long messages, we consider a Fixed Reservation scheme to share a communication channel on a store-and-forward basis. Under this scheme, the channel time is divided into two fixed portions, one portion for sources to make reservations and the other portion for actual message transmission. A source that has a message to transmit uses the reservation portion of the channel to make reservation. When reservation for the message has been successfully made, the message is then scheduled for future transmission. The source will transmit the message at the channel transmission rate in its reserved slots. Since interruptions for more important messages are acceptable and practical in data transmission applications, we analyze the delay-throughput performance of Fixed Reservation schemes using a preemptive priority discipline.

## 1.2 History

Years ago, channel multi-accessing was accomplished by circuit-switching methods, such as Fixed-Assignment FDMA schemes. These

methods were adequate for a long time. In recent years, data transmission applications have substantially expanded. Consequently, not only has the amount of communication needs grown, but also the diversity of message characteristics and complexity of source requirements have developed. Conventional communication techniques are no longer adequate for the efficient sharing of communication channels. With the decrease in cost of processing and processor, attention has shifted to packet-switching forms of multiple access. A transmitter formats messages into packets of constant length and then transmits them over the channel. This can be carried out in real-time transmission and store-and-forward transmission. Most of the work that has been done on this form of multiple access is as follows.

Abramson [7] and [11] studied the Aloha system (Random Access). It is unstable and it has a low maximum throughput of  $1/2e$ .

Roberts [8] improved the maximum throughput of the channel to  $1/e$  by slotting the channel time and synchronizing the sources so that the packets arrive in phase (Slotted Random Access).

Metcalfe [12] and [13] tried to stabilize the Aloha and the Slotted Aloha systems by two methods. First, he introduced blocking, that is, a source may not generate another packet until the present packet has been transmitted successfully. Second, he varied the transmission probabilities.

Kleinrock and Lam [14] and [15] extended Metcalfe's work. They showed that even by blocking, if the population of sources is infinite and independent, the system is unstable. If the population of sources is finite, then the system is stable. However, due to

traffic fluctuations, the system sways between two points, one with a small number of retransmitted packets and another with a large number of retransmitted packets.

Rubin [10] introduced Group Random Access schemes. These schemes have been described in the previous section. He also introduced blocking in order to stabilize the channel.

Capetanakis [9] proposed Tree Search schemes. Under a Tree Search scheme, sources are arranged to form a tree structure. Sources are allowed to transmit only when they are given the transmission rights. Transmission rights are given to sources by following a search pattern along the branches of the tree.

So far, under the schemes we have discussed, if more than one packet is transmitted simultaneously, then a collision occurs and all the packets are destroyed. However, Roberts [8] showed that this need not be true because FM receivers can track the strongest of many signals if the next strongest is down by 1.5 to 3 db. In a ground radio system, he increased the channel maximum throughput to 0.6 by taking advantage of this FM capture.

There are other schemes which can improve the channel throughput by requiring each source to obtain more information about the other sources. However, these schemes are efficient only if the round trip propagation delay of the channel is small. Kleinrock and Tobagi [16], [17], [18] and [19] examined CSMA (Carrier Sense Multiple Access). In CSMA, a source with a packet to transmit first listens for the carrier of other sources to determine if the channel is busy. If it is busy then the source holds the packet. Otherwise, it transmits the packet



with certain probability. Unavoidable collisions are resolved by the Aloha technique.

Kleinrock and Yemini [20] suggested Urn schemes under which sources that have packets to transmit are required to send signals in mini slots. Hence, every source can acquire information about the other sources by listening to these mini slots. Then based on this information, transmission rights in each slot are assigned among the sources by following a certain rule.

In addition to these schemes, a number of reservation schemes have been proposed. Crowther et al [21] introduced Reservation Aloha schemes under which the channel slots are grouped into frames that are at least one propagation delay long. A source that has successfully used a slot retains transmission right in that slot in the following frame. Unused slots can be captured by any source through the Slotted Aloha technique.

Roberts [22] suggested Interleaved Reservation Aloha schemes under which the channel is divided into two states, Reservation and Aloha. On the Reservation state, the sources try to reserve the Aloha state by the Slotted Aloha technique.

Binder [23] introduced reservation schemes under which the slots are grouped into frames such that each source is allocated a slot in every frame. A slot which is not used by its owner is available to the other sources on a Round Robin basis.

Rubin [24] offered Fixed Reservation schemes under which the channel frame is divided into two portions, one portion for sending reservation information and the other portion for actual message

transmission. By changing the frame duration and the ratio of the two portions, Dynamic and Fixed Reservation schemes are obtained which are capable of adapting to traffic flow fluctuations.

Instead of fixing the channel frame into two portions, Rubin [24] and [25] developed the Asynchronous Reservation Demand-Assignment schemes by establishing reservation slots dynamically according to observed service demands and queue sizes.

### 1.3 The System Models

In the following chapters, we will describe several situations by various system models. Each system is composed of a communication channel and the sources that share it. To completely describe and analyze a system, it is necessary to specify the characteristics of the channel and the sources, the stochastic properties of the messages arriving at the sources and the access-control scheme that allocates the capacity of the channel to the sources. To accomplish this, a channel model, a source model or (and) message model and an access-control scheme model are set up.

### 1.4 Outline of the Dissertation

In Chapter II, Demand-Assignment TDMA schemes used to grant channel access to real-time sources which require transmission at a specific source information rate are discussed. A Cutoff Priority discipline is employed to give advantage to priority messages.

In Chapter III, Demand-Assignment TDMA schemes for real-time sources which require transmission at different source information rates are presented. An overall message delay performance measure is

introduced and used as an objective function to optimally allocate the capacity of the channel among the sources.

In Chapter IV, Fixed-Assignment TDMA and conflict resolving random access schemes are combined to form a class of hybrid access-control schemes. Two versions of the Tree Search technique are combined with Fixed-Assignment TDMA schemes to form two families of hybrid access-control schemes. A technique is developed to obtain numerical results for the delay-throughput performance of these hybrid schemes through recursive calculations. Two cases are studied. In one case, the sources are assumed to have limited buffer capacities, and in the other case, unlimited buffer capacities. To demonstrate the adaptability of both families of schemes over the throughput range, global delay-throughput performance curves of the members of each family are obtained. The performance of the two families is then compared by numerical examples.

In Chapter V, we study the performance of Fixed Reservation schemes. A mathematical technique (Theorem 5.4) is employed to modify Rubin's results to obtain tighter bounds on the average message delay. In some special cases, these bounds converge to give exact results. The technique is further used to study the effect of employing a preemptive priority discipline under Fixed Reservation schemes. As a special case, numerical results are obtained for the mean delay of a preemptive priority Fixed-Assignment TDMA scheme.

In Chapter VI, conclusions and suggestions for future research are presented.

CHAPTER II  
DEMAND-ASSIGNMENT TDMA SCHEMES WITH CUTOFF PRIORITY  
FOR REAL-TIME SOURCES

2.1 Introduction

In this chapter, we consider the sharing of a communication channel by sources which require real-time transmission at a specific information rate. These sources require real-time transmission. Hence, once access is assigned, a source will transmit at a specific information rate. Since message arrival times and message holding times are a priori unknown, to efficiently utilize the channel capacity and yield acceptable average message delay, it is necessary to design an access-control scheme to allocate channel capacity among the sources. We consider Demand-Assignment TDMA schemes.

In practical situations, some messages are more important than others and they demand higher priority than others when requesting access. On the other hand, it is undesirable to interrupt or preempt a source once access is assigned to it. Our goal is to design an access-control scheme which will give advantage to the important messages and yet won't interrupt any source which has been assigned access.

2.2 The System Models

In the following, we present the channel model, the Demand-Assignment TDMA schemes and the message model that will be used to describe the systems we are going to study in this chapter.

### The Channel Model

The channel is time-slotted, that is, its channel time is divided into equal durations called slots. These slots are further arranged to form consecutive blocks of slots called frames. Each frame is made up of a fixed number of consecutive slots. When a source is granted permission to transmit, it will transmit in a sequence of slots of fixed frame position in successive frames. For example, if a source is assigned the first slot, then it will transmit in the first slot of every frame until it finishes and relinquishes the slot. (This is illustrated in Figure 2.1.) Therefore, each source can transmit information at a fixed information rate. The slot length and the frame duration are designed in such a way that this fixed information rate matches that required by the sources. Set  $K$  to be the number of slots per frame. Then the channel can support  $K$  sources simultaneously.

### Demand-Assignment TDMA Schemes

If the message traffic arriving at the sources is steady, then we can employ a Fixed-Assignment TDMA scheme by dedicating one slot in each frame to every source. However, if the message traffic arriving at the sources is bursty (statistically fluctuating), a Fixed-Assignment TDMA scheme will yield low channel utilization. In that case, to efficiently allocate channel slots among the sources, we employ a Demand-Assignment TDMA scheme.

Under a Demand-Assignment TDMA scheme, sources which have messages to transmit are required to send requests. A source can

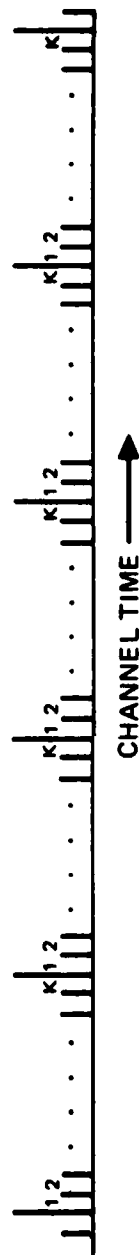


Figure 2.1. TDMA Time Frame Structure.

transmit only when its request is granted. Two procedures for sending requests are described. One is central control and the other is distributed control.

Under central control, a central controller is established. Sources that have messages to transmit send requests to the central controller. Upon receiving these requests, the controller assigns available slots to the sources. If there is no available slot, then the requests are lost or queued in a request buffer where they will receive permissions to transmit when there are available slots. When the transmission of a message is finished, the slot is relinquished by sending a signal at the end of the message or by sending a separate finish signal to the controller.

Under distributed control, sources that have messages to transmit send their requests to all the sources through a broadcast channel. Hence, every source knows which are the sources requesting slots. Based on this information, the requests are granted, lost, or queued and granted permissions later according to an agreed upon policy.

In both cases, a channel is required for sending control information. This can be done by establishing a separate channel or by assigning a portion of the channel under consideration.

#### The Message Model

Sources which have messages to transmit send requests for slots, one request for one message. We assume the message arrival process to be Poisson with average arrival rate  $\lambda$  (messages per second). The holding times of the messages are assumed to be independent and identically

distributed with mean  $\mu^{-1}$  (seconds).

### 2.3 The Performance of Demand-Assignment TDMA Schemes

In order to store requests when immediate access cannot be assigned, a request buffer is established. This buffer may have limited or unlimited capacity. Here, we present the results of two systems. In one system, the request buffer is assumed to have unlimited buffer capacity and in the other, limited buffer capacity  $N$ . In the latter system, upon the arrival of a new request, if the request buffer is full, the request is regarded to be lost.

#### Request Buffer with Unlimited Capacity

If the buffer used to store requests is assumed to have unlimited capacity, then the system can be approximated by an M/M/K queueing system (see Appendix and [26]). The average waiting time of a request is given by

$$E[W] = P(K) \left\{ K\mu(1-\rho)^2 \right\}^{-1}, \text{ for } \rho < 1, \quad (2.1)$$

where

$$\rho = \lambda(K\mu)^{-1}, \quad (2.2)$$

$$P(K) = P(0) \frac{1}{K!} \left( \frac{\lambda}{\mu} \right)^K, \quad (2.3)$$

and

$$P(0) = \left\{ 1 + \frac{\lambda}{\mu} + \dots + \frac{1}{(K-1)!} \left( \frac{\lambda}{\mu} \right)^{K-1} + \frac{1}{(1-\rho)} \frac{1}{K!} \left( \frac{\lambda}{\mu} \right)^K \right\}^{-1} \quad (2.4)$$

$$E[W] = \infty, \text{ for } \rho \geq 1. \quad (2.5)$$



Since the request buffer has unlimited capacity, the loss probability of a request is 0.

#### Request Buffer with Limited Capacity

If the buffer used to store request is assumed to have limited capacity,  $N$  (requests), then the system can be approximated by an  $M/M/K/N$  queueing system. The average waiting time of an accepted request is given by

$$E[W] = P(K) \left( 1 + N\rho^{N+1} - (N+1)\rho^N \right) \left\{ K\mu(1-\rho)^2 \right\}^{-1}, \quad (2.6)$$

where

$$\rho = \lambda(K\mu)^{-1}, \quad (2.7)$$

$$P(k) = \begin{cases} P(0) \frac{1}{k!} \left( \frac{\lambda}{\mu} \right)^k, & 0 \leq k \leq K, \\ P(0) \frac{1}{K!} \left( \frac{\lambda}{\mu} \right)^K \rho^{k-K}, & K < k \leq K + N \end{cases} \quad (2.8)$$

and

$$P(0) = \left\{ 1 + \frac{\lambda}{\mu} + \dots + \frac{1}{(K-1)!} \left( \frac{\lambda}{\mu} \right)^{K-1} + \frac{1 - \rho^{N+1}}{(1-\rho)} \frac{1}{K!} \left( \frac{\lambda}{\mu} \right)^K \right\}^{-1}. \quad (2.9)$$

The loss probability of a request is

$$\ell = P(K)\rho^N. \quad (2.10)$$

#### 2.4 The Performance of Demand-Assignment TDMA Schemes with Cutoff

##### Priority

Among the messages that require transmission over the channel, some of them are more important than others. We call these messages

priority messages and the others ordinary messages. It is desirable to give advantage to these priority messages and yet not interrupt any ordinary messages which are being transmitted. For this purpose, we use a Cutoff Priority discipline. It is defined as follows. Recall that there are  $K$  slots in every frame. Let  $K_1$  be a number such that  $K_1 \leq K$ . At any time, if the number of occupied slots is greater than or equal to  $K_1$ , then no ordinary requests are granted, even if there are vacant slots. However, ordinary messages are not preempted by priority messages, that is, if a slot has been assigned to an ordinary message, the latter can keep the slot until it is finished, even if there are times during its occupation of the slot when the number of occupied slots is equal to or greater than  $K_1$ .  $K_1$  is defined to be the cutoff threshold.

A request, either priority or ordinary, which is not granted a slot can be put into a queue or regarded as lost. If it is queued, we assume the request buffer has unlimited capacity. There are two types of requests (priority messages and ordinary messages) and there are two ways of treating each type of requests when immediate slots cannot be assigned. Hence, there are four system combinations. (This is described by Table 2.1.)

Before we present the results of these four cases, we make the following assumptions.

#### Assumption 2.1

The arrival process of the priority messages is assumed to be Poisson with average arrival rate  $\lambda_1$  (messages per second) and the

Assumption made when immediate  
slot(s) cannot be granted

CASE NUMBER	PRIORITY REQUEST	ORDINARY REQUEST
I	lost	lost
II	queued	lost
III	lost	queued
IV	queued	queued

Table 2.1. Four ways of treating Requests when  
immediate slot(s) cannot be granted.

arrival process of the ordinary messages is also assumed to be Poisson with average arrival rate  $\lambda_2$  (messages per second). Both processes are assumed to be statistically independent.

#### Assumption 2.2

The holding times of both priority and ordinary messages are assumed to be independent and identically distributed according to an exponential distribution with mean  $\mu^{-1}$  (seconds).

We also define the following terms:

- 1)  $\lambda = \lambda_1 + \lambda_2$ .
- 2)  $\sigma_i = \lambda_i / \mu$ ,  $i = 1, 2$ .
- 3)  $\sigma = \lambda / \mu$ .
- 4)  $\ell_1$  - the loss probability of a priority request.
- 5)  $\ell_2$  - the loss probability of an ordinary request.
- 6)  $W_1$  - the waiting time of an accepted priority request.
- 7)  $W_2$  - the waiting time of an accepted ordinary request.

In the following, we will present results of the four cases without going into details. The derivation of the results are based on the analysis of the joint queue size of both types of requests which forms an underlining Markov Chain (see [27]).

#### Case I

In this system, both types of requests are lost when no immediate slot can be assigned. We have

$$l_1 = \hat{e} \sigma_1^{K_1} \sigma_1^{K-K_1} (K!)^{-1}, \quad (2.11)$$

and

$$l_2 = \hat{e} \left( \frac{\sigma}{\sigma_1} \right)^{K_1} \sum_{k=K_1}^K \sigma_1^k (k!)^{-1}, \quad (2.12)$$

where

$$\frac{1}{\hat{e}} = \sum_{k=0}^{K_1} \sigma^k (k!)^{-1} + \left( \frac{\sigma}{\sigma_1} \right)^{K_1} \sum_{k=K_1+1}^K \sigma_1^k (k!)^{-1}. \quad (2.13)$$

Obviously,

$$E[W_1] = E[W_2] = 0. \quad (2.14)$$

#### Case II

In this system, the priority requests are queued and the ordinary requests are lost when no immediate slot can be assigned. We have

$$P(W_1 > 0) = \hat{e} \sigma^{K_1} \sigma_1^{K-K_1} \left\{ (K-1)! (K-\sigma_1) \right\}^{-1}, \quad (2.15)$$

$$\begin{aligned} E[W_1] &= \sigma^{K_1} \sigma_1^{K-K_1} [K^2 - (K-1)\sigma_1] \hat{e} \\ &\quad \cdot \left\{ (K-1)! (K-\sigma_1)^2 \right\}^{-1}, \end{aligned} \quad (2.16)$$

and

$$\begin{aligned} l_2 &= \hat{e} \left\{ \left( \frac{\sigma}{\sigma_1} \right)^{K_1} \sum_{k=K_1}^K \sigma_1^k (k!)^{-1} \right. \\ &\quad \left. + \sigma^{K_1} \sigma_1^{K-K_1+1} (K! (K-\sigma_1))^{-1} \right\}, \quad \text{for } \sigma_1 < K. \end{aligned} \quad (2.17)$$

where

$$\begin{aligned} \frac{1}{\hat{e}} = & \sum_{k=0}^{K_1} \sigma_1^k (k!)^{-1} + \left( \frac{\sigma}{\sigma_1} \right)^{K_1} \sum_{k=K_1+1}^K \sigma_1^k (k!)^{-1} \\ & + \sigma_1^{K_1} \sigma_1^{K-K_1+1} (K!(K-\sigma_1))^{-1}. \end{aligned} \quad (2.18)$$

Obviously,

$$P(W_1 > 0) = \ell_2 = 1, \quad \text{for } \sigma_1 \geq K. \quad (2.19)$$

### Case III

In this system, the priority requests are lost and the ordinary requests are queued when no immediate slot can be assigned. We have

$$\begin{aligned} \ell_1 = & \frac{\sigma_1^{K_1}}{\sigma_2(K_1-1)!} \left\{ \frac{\sigma_2 - K_1}{\sigma_2} \frac{K!}{K_1!} \sigma_1^{K_1-K} \right. \\ & \left. + K! \sigma_1^{-K} \sum_{k=K_1+1}^K \sigma_1^k (k!)^{-1} \right\}^{-1} \hat{e}, \end{aligned} \quad (2.20)$$

and

$$P(W_2 > 0) = \ell_1 K! \sigma_1^{-K} \sum_{k=K_1}^K \sigma_1^k (k!)^{-1}, \quad (2.21)$$

where

$$\begin{aligned} \frac{1}{\hat{e}} = & \sum_{k=0}^{K_1-1} \frac{\sigma_1^k}{k!} - \frac{\sigma_1^{K_1}}{(K_1-1)! \sigma_2} \left\{ \frac{(\sigma_2 - K_1)}{\sigma_2} \sigma_1^{K_1} (K_1!)^{-1} \right. \\ & \left. + \sum_{k=K_1+1}^K \sigma_1^k (k!)^{-1} \right\}^{-1} \sum_{k=K_1}^K \sigma_1^k (k!)^{-1}. \end{aligned} \quad (2.22)$$

#### Case IV

In this system, requests of both types are queued if no immediate slot can be assigned. No useful result of simple form can be obtained.

In all the above cases, if there are more than two types of requests, hence, more than one cutoff threshold, the results would be much more complicated. However, in Case IV, if the cutoff thresholds for all types of requests are set to be  $K$ , then it can be approximated by a non-preemptive multiple server queueing system. This has been studied by Cobham (see [27]). For the  $j$  Class of requests, we have

$$E[W_j] = \frac{\pi}{K\mu} \left\{ \left( 1 - \frac{1}{K\mu} \sum_{i=1}^{j-1} \lambda_i \right) \cdot \left( 1 - \frac{1}{K\mu} \sum_{i=1}^j \lambda_i \right) \right\}^{-1}, \quad (2.23)$$

where

$$\pi = \sigma^K \left\{ K! \left( 1 - \frac{\lambda}{K\mu} \right) \cdot \left[ \sum_{j=0}^{K-1} \frac{\sigma^j}{j!} + \sum_{j=K}^{\infty} \sigma^j (K! K^{j-K})^{-1} \right] \right\}^{-1}. \quad (2.24)$$

#### Numerical Results

To demonstrate the performance of Demand-Assignment TDMA schemes with Cutoff Priority discipline, we consider a system in which  $K = 10$ ,  $\lambda_1 = 0.05$  and  $\mu = 0.1$ . Figures 2.2-2.4 plot the performance curves of the system when  $0.1 \leq \lambda_2 \leq 0.4$ .

Figure 2.2 plots the loss probability of a Type 1 request,  $\ell_1$ , versus the loss probability of a Type 2 request,  $\ell_2$ , under different

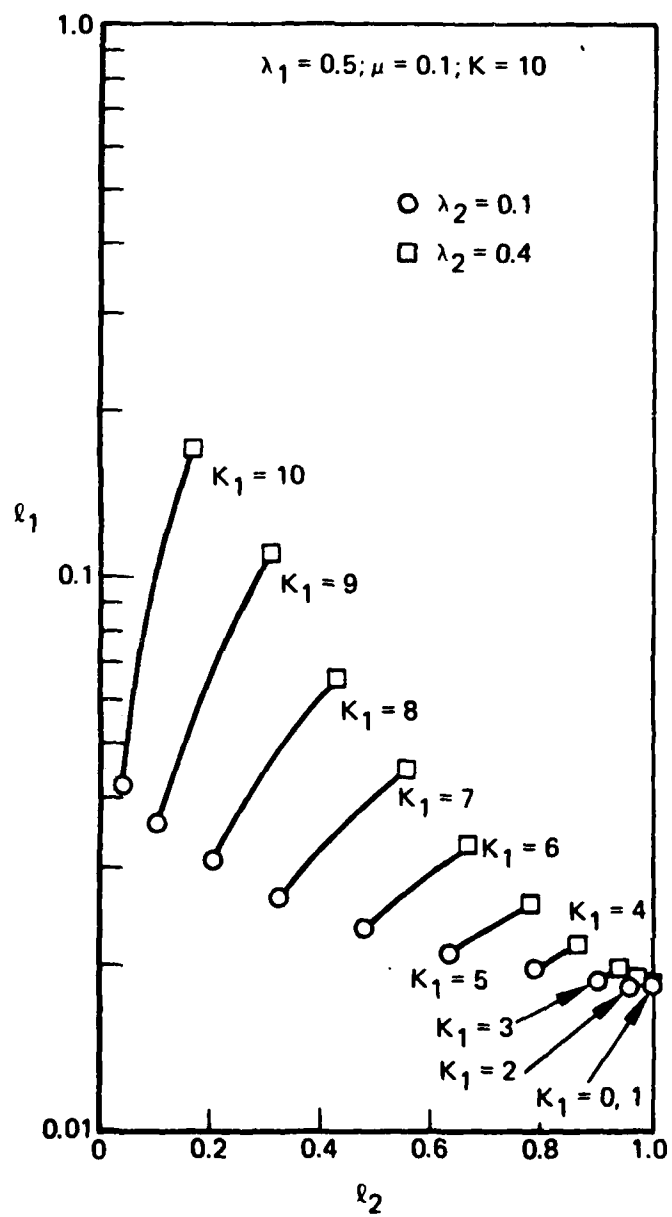


Figure 2.2.  $l_1$  vs  $l_2$  Curves of Case I.



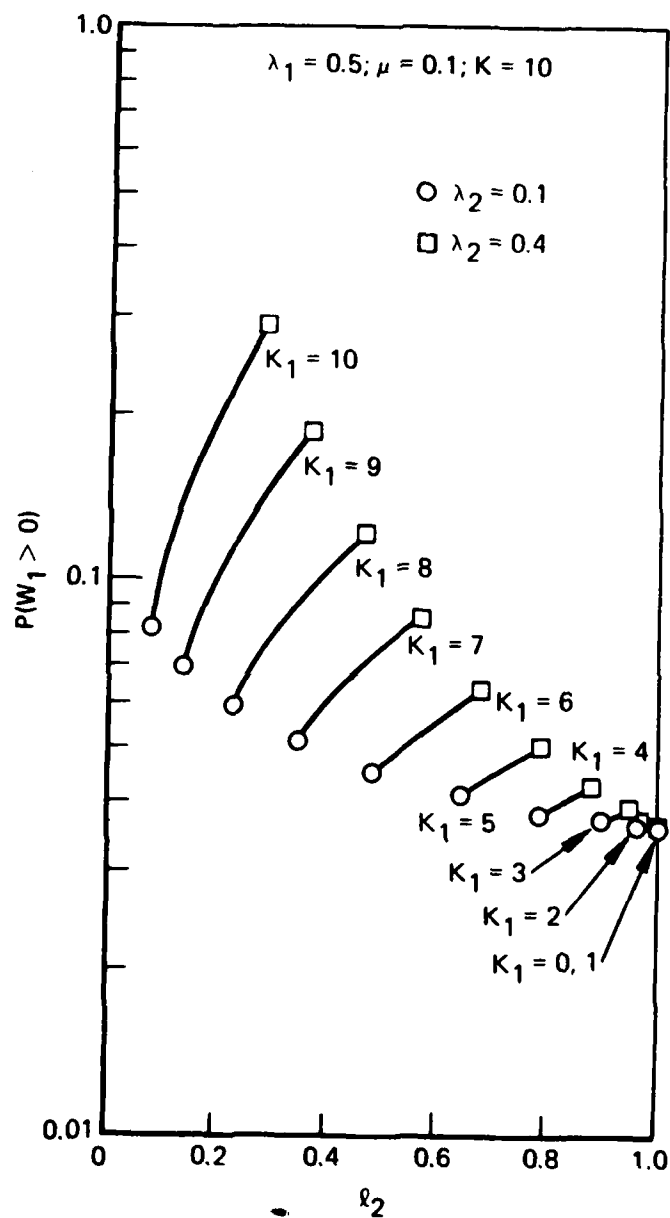


Figure 2.3.  $P(W_1 > 0)$  vs  $l_2$  Curves of Case II.

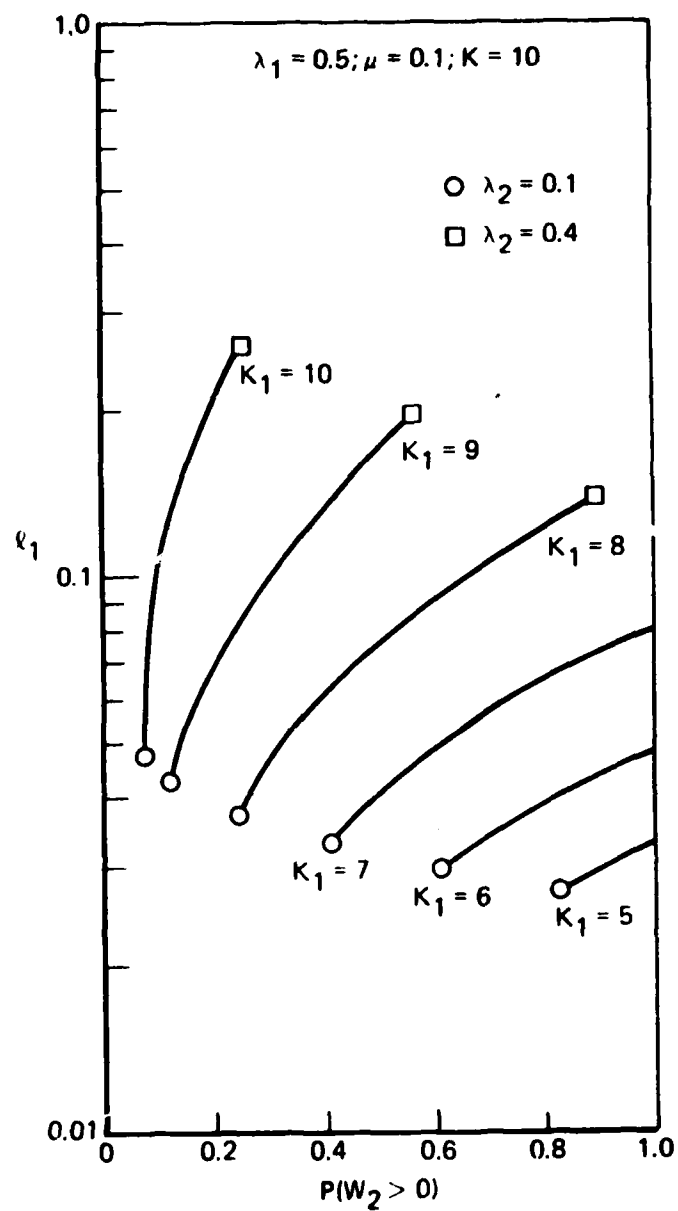


Figure 2.4.  $\ell_1$  vs  $P(W_2 > 0)$  Curves of Case III.

values of  $K_1$ , when the requests of the system are treated according to the policy described in Case I. For any fixed value of  $\lambda_2$  (in particular,  $\lambda_2 = 0.1$  and  $\lambda_2 = 0.4$  in the figure), the smaller is  $K_1$ , the smaller is the number of slots available for Type 2 requests and the more is the advantage given to Type 1 requests. Hence, a threshold of lower  $K_1$  value yields lower  $\ell_1$  value and higher  $\ell_2$  value than a threshold of higher  $K_1$  value. For any  $K_1$  value, when  $\lambda_2$  is increased, the loss probability of both types of requests are increased. However, a threshold of higher  $K_1$  value means more slots available for Type 2 requests and therefore more influence on the total traffic by Type 2 traffic than a threshold of lower  $K_1$  value. Hence, a curve of higher  $K_1$  value is steeper than a curve of lower  $K_1$  value.

Figure 2.3 plots the probability of waiting of a Type 1 request versus the loss probability of a Type 2 request,  $\ell_2$ , under different values of  $K_1$ , when the requests of the system are treated according to the policy described in Case II. This set of curves are similar to those in Figure 2.2. However, in this set of curves, Type 1 requests are queued, not lost, when no immediate slot can be assigned. Hence, this set of curves are higher than those in Figure 2.2.

Figure 2.4 plots the loss probability of a Type 1 request,  $\ell_1$ , versus the probability of waiting of a Type 2 request, under different values of  $K_1$ , when the requests of the system are treated according to the policy described in Case III. This set of curves are similar to those in Figures 2.2 and 2.3. However, in this case, Type 2 requests are queued, not lost, when no immediate slot can be assigned.

Hence, this set of curves are lower than those in Figure 2.3 but shifted to the right. Also, for  $K_1 < 8$ , the system becomes unstable with respect to Type 2 requests, that is,  $E[W_2] = \infty$ , as the value of  $\lambda_2$  is increased to a rate which is smaller than 0.4.

When the requests of the system are treated according to the policy described in Case IV, both types of requests are queued, not lost, when no immediate slot can be assigned. In this case, when  $K_1 = K$ , the probability of waiting of either a Type 1 or Type 2 request is 0.01 if  $\lambda_2 = 0.01$  and is 0.669 if  $\lambda_2 = 0.04$ . This is higher and more shifted to the right than any of the previous cases with the same  $K_1$  value.

CHAPTER III  
DEMAND-ASSIGNMENT TDMA SCHEMES WITH FIXED BOUNDARIES  
FOR MULTIPLE-RATE REAL-TIME SOURCES

3.1 Introduction

In the last chapter, we discussed the sharing of a time-slotted communication channel by a number of sources which require real-time transmission at a specific information rate. In this chapter, we consider systems in which sources require real-time transmission at different information rates. In our analysis, we consider the case in which sources transmit at two multiple information rates. However, the method can be extended, inducing algebraic complication, to cases in which sources transmit at more than two multiple information rates. Demand-Assignment TDMA schemes with fixed boundaries are used to allocate channel slots to sources which have messages to transmit. To quantitatively assess these schemes, a system performance measure is introduced. Subsequently, a procedure is developed to find the optimal Demand-Assignment TDMA frame structure. Finally, a numerical example is presented. In this example, the performance curves of the Demand-Assignment TDMA schemes are plotted. The procedure of finding the optimal TDMA frame structure is also demonstrated.

3.2 The System Models

In the following, we will present the source model, the channel model, the Demand-Assignment TDMA schemes and the message model which will be used to describe the systems we are going to study in this chapter.

### The Source Model

The sources which share the channel require real-time transmission at multiple information rates. Consequently, they can be divided into two types according to their information rates. Type 1 sources transmit at information rate  $R_1$  (bits per second). Type 2 sources transmit at information rate  $R_2$  (bits per second). We assume that  $R_1$  is an integer multiple of  $R_2$ . Hence,  $R_1 = MR_2$ , where  $M$  is an integer.

### The Channel Model

The channel time is divided into slots of equal durations. The slots are further arranged to form consecutive blocks of equal lengths called frames. Since the two types of sources require real-time transmission at  $R_1$  and  $R_2$ , respectively, the slot length and the frame length are designed in such a way that to support a Type 2 transmission would require a sequence of slots (one slot of fixed frame position per frame) in successive frames and to support a Type 1 transmission would require a sequence of groups ( $M$  contiguous slots of fixed frame positions per frame) of slots in successive frames.

### Demand-Assignment TDMA Schemes with Fixed Boundaries

If the message traffic is steady, a Fixed-Assignment TDMA scheme can be employed by assigning a portion of the channel frame to every source. However, if the message traffic is bursty, a Fixed-Assignment TDMA scheme will yield low channel utilization. It is then more efficient to employ a Demand-Assignment TDMA scheme. Under a Demand-Assignment TDMA scheme, the sources which have messages to transmit are required to send request information before they can be

granted permission to transmit. The method of sending request information has been discussed in the last chapter and is not to be repeated here. We are interested in how to allocate the channel slots among the sources which have sent requests.

One simple strategy is described as follows. The channel frame is divided into two fixed portions, Portion 1 for Type 1 sources and Portion 2 for Type 2 sources. The two portions of the channel operate independently. Type 1 sources can only use the slots in Portion 1 and Type 2 sources can only use the slots in Portion 2. If there are  $MK_1$  slots and  $K_2$  slots in Portion 1 and Portion 2, respectively, then the channel can support  $K_1$  Type 1 sources and  $K_2$  Type 2 sources simultaneously.

#### The Message Model

Messages arriving at Type 1 sources are called Type 1 messages and messages arriving at Type 2 sources are called Type 2 messages. The arrivals of the two types of messages are assumed to be statistically independent. The arrival processes of messages at Type  $j$ ,  $j = 1, 2$ , sources are assumed to be Poisson with parameters  $\lambda_j$  (messages per second),  $j = 1, 2$ . The holding times of the two types of messages are statistically independent. The holding times of Type  $j$ ,  $j = 1, 2$ , messages are independent and identically distributed, according to an exponential distribution with mean  $\mu_j^{-1}$  (seconds). In order to employ an M/M/K queueing system in our analysis, the means of the holding times of both types of messages are assumed to be much longer than the channel frame time (see Appendix).

### 3.3 The Performance of Demand-Assignment TDMA Schemes

As described in the last chapter, a request buffer is established to store requests when immediate slot(s) cannot be assigned. Consequently, we present the results of two cases. In one case, the request buffer is assumed to have unlimited capacity and in the other, limited capacity  $N$ . Since the two portions of the frame operate independently, a system with unlimited or limited buffer capacities for either type of requests can be described by these two cases.

#### Request Buffer with Unlimited Capacity

If the buffer used to store Type  $j$  requests,  $j = 1, 2$  is assumed to have unlimited capacity, then the operation of Portion  $j$ ,  $j = 1, 2$ , can be approximated by an  $M/M/K_j$  queueing system (see Appendix and [26]). The average waiting time of a Type  $j$  request,  $j = 1, 2$  is given by

$$\begin{aligned}
 & E[W_j(K_j, \lambda_j)] \\
 &= P_{X_j}(K_j) \left\{ K_j \mu_j \left( 1 - \frac{\lambda_j}{K_j \mu_j} \right)^{K_j-1} \right\}, \\
 & \text{for } \lambda_j < K_j \mu_j,
 \end{aligned} \tag{3.1}$$

where

$$P_{X_j}(K_j) = P_{X_j}(0) \frac{1}{K_j!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j}, \tag{3.2}$$



and

$$P_{X_j}(0) = \left\{ 1 + \frac{\lambda_j}{\mu_j} + \dots + \frac{1}{(K_j-1)!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j-1} + \left( 1 - \frac{\lambda_j}{K_j \mu_j} \right)^{-1} \frac{1}{K_j!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j} \right\}^{-1} \quad (3.3)$$

$$E[W_j(K_j, \lambda_j)] = \infty, \text{ for } \lambda_j \geq K_j \mu_j \quad (3.4)$$

#### Request Buffer with Limited Capacity

If the buffer used to store Type  $j$  requests is assumed to have limited capacity  $N_j$ ,  $j = 1, 2$ , then the operations of Portion  $j$ ,  $j = 1, 2$ , can be approximated by an  $M/M/K_j/N_j$  queueing system. The average waiting time of an accepted Type  $j$  request,  $j = 1, 2$ , is given by

$$\begin{aligned} E[W_j(K_j, \lambda_j)] &= P_{X_j}(K_j) \left\{ 1 + N_j \left( \frac{\lambda_j}{K_j \mu_j} \right)^{N_j+1} - (N_j+1) \left( \frac{\lambda_j}{K_j \mu_j} \right)^{N_j} \right\} \left\{ K_j \mu_j \left( 1 - \frac{\lambda_j}{K_j \mu_j} \right)^2 \right\}^{-1}, \end{aligned} \quad (3.5)$$

where

$$P_{X_j}(K_j) = P_{X_j}(0) \frac{1}{K_j!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j}, \quad (3.6)$$

and

$$P_{X_j}(0) = \left\{ 1 + \frac{\lambda_j}{\mu_j} + \dots + \frac{1}{(K_j-1)!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j-1} + \left[ 1 - \left( \frac{\lambda_j}{K_j \mu_j} \right)^{N_j+1} \right] \left( 1 - \frac{\lambda_j}{K_j \mu_j} \right)^{-1} \frac{1}{K_j!} \left( \frac{\lambda_j}{\mu_j} \right)^{K_j} \right\}^{-1} \quad (3.7)$$

The loss probability of a Type  $j$  request,  $j = 1, 2$ , is

$$l_j = P_{X_j}(K_j) \left( \frac{\lambda_j}{K_j \mu_j} \right)^{N_j} \quad (3.8)$$

### 3.4 Optimal Allocation of Slots

Under a Demand-Assignment TDMA scheme, the channel frame is divided into two portions,  $MK_1$  slots in Portion 1 and  $K_2$  slots in Portion 2,  $K_2 = K - MK_1$ . Our goal is to find the optimal  $K_1$ , consequently,  $K_2$ , so that a prescribed level of performance is guaranteed. We will study the system in which the request buffers for both types of requests are assumed to have unlimited capacities.

One appropriate performance measure of the system is the maximum normalized average waiting time, denoted by  $W$ , of the requests. This is defined as

$$W(K_1, K_2, \lambda_1, \lambda_2) = \text{Max}[\tilde{W}_1(K_1, \lambda_1), \tilde{W}_2(K_2, \lambda_2)] ,$$

where

$$\tilde{W}_j(K_j, \lambda_j) = E[W_j(K_j, \lambda_j)] / \mu_j^{-1}, \quad j = 1, 2.$$

Therefore,  $\tilde{W}_j(K_j, \lambda_j)$  denotes the ratio of the average waiting time of

a Type  $j$  request,  $j = 1, 2$ , to the average holding time of a Type  $j$  message,  $j = 1, 2$ . According to this system performance measure, the optimal allocation of slots, denoted by  $MK_1^*$  and  $K_2^*$ , yields the minimum value of  $W(\cdot)$ , denoted by  $W^*(\cdot)$ . Therefore,

$$W^*(K_1^*, K_2^*, \lambda_1, \lambda_2) = \min_{K_1} W(K_1, K - MK_1, \lambda_1, \lambda_2).$$

In order to find  $K_1^*$  and  $K_2^*$ , given  $\lambda_1$  and  $\lambda_2$ , we make the following definitions:

$\Omega(K_1)$  - for each given  $K_1$ , the region of average message arrival rates,  $(\lambda_1, \lambda_2)$ , which yields a finite normalized average waiting time. Therefore,

$$\Omega(K_1) = \{(\lambda_1, \lambda_2), \lambda_1 \geq 0, \lambda_2 \geq 0: \tilde{W}_1(K_1, \lambda_1) < \infty$$

$$\text{and } \tilde{W}_2(K - MK_1, \lambda_2) < \infty\}.$$

$\Psi(K_1)$  - for each given  $K_1$ , the region of average message arrival rates  $(\lambda_1, \lambda_2)$ , which yields a finite normalized average waiting time in which  $K_1^* = K_1$ . Therefore,

$$\Psi(K_1) = \{(\lambda_1, \lambda_2): (\lambda_1, \lambda_2) \in \Omega(K_1) \text{ and } K_1^* = K_1\}.$$

Let  $[X]$  denote the largest integer less than  $X$ . We introduce the following theorem which yields a procedure for calculating  $K_1^*$  given  $(\lambda_1, \lambda_2)$ .

### Theorem 3.1

$$\text{Partition the region } \Omega = \bigcup_{i=1}^{[K/M]} \Omega(i)$$

into the disjoint region  $A(K_1)$ ,  $1 \leq K_1 \leq [K/M]$ ,

$$\Omega = \bigcup_{K_1=1}^{[K/M]} A(K_1) ,$$

where  $A(K_1) \subset \Omega(K_1)$  and is further contained between boundaries  $B(K_1-1)$  and  $B(K_1)$ , where

$$B(0) = \{(\lambda_1, \lambda_2) : \lambda_1 = 0\} ,$$

$$B(K_1) = \{(\lambda_1, \lambda_2) : \tilde{W}(K_1, \lambda_1) = \tilde{W}_2(K-M(K_1+1), \lambda_2) < \infty\} ,$$

$$1 \leq K_1 \leq [K/M] - 1 ,$$

$$B([K/M]) = \{(\lambda_1, \lambda_2) : \lambda_2 = 0\} .$$

Then,

$$\Psi(K_1) = A(K_1) . \quad (3.9)$$

#### Proof

Let

$$C(K_1) = \{\Omega(K_1-1) \cap \Omega(K_1)\} \\ \cup \{\Omega(K_1) \cap \Omega(K_1+1)\} .$$

By construction, any point  $(\lambda_1, \lambda_2)$  which belongs to the region  $\Omega(K_1) - C(K_1)$  satisfies the following equations.

$$\tilde{W}_1(K_1-1, \lambda_1) = \infty , \quad (3.10)$$

$$\tilde{W}_2(K-M(K_1+1), \lambda_2) = \infty , \quad (3.11)$$

$$\tilde{W}_1(K_1, \lambda_1) < \infty , \quad (3.12)$$

and

$$\tilde{W}_2(K-MK_1, \lambda_2) < \infty . \quad (3.13)$$

Consequently, in this region

$$\tilde{W}_1(K_1-1, \lambda_1) > \tilde{W}_1(K_1-1, \lambda_1) = \infty, \text{ for } 1 < i < K_1. \quad (3.14)$$

and

$$\begin{aligned} \tilde{W}_2(K-M(K_1+1), \lambda_2) > \tilde{W}_2(K-M(K_1+1), \lambda_2) = \infty, \\ \text{for } 1 < i \leq [K/M] - K_1. \end{aligned} \quad (3.15)$$

Therefore, for any point  $(\lambda_1, \lambda_2)$  in the region  $\Omega(K_1) - C(K_1)$ ,  $K_1^* = K_1$ .

Now, we consider the region  $C(K_1)$ . First, we examine the Boundary  $B(K_1-1)$ . From the construction of  $C(K_1)$ , we have

$$\begin{aligned} \{(\lambda_1, \lambda_2): 0 \leq \lambda_1 < (K_1-1)\mu_1, \\ 0 \leq \lambda_2 < (K-MK_1)\mu_2\} \subset C(K_1). \end{aligned} \quad (3.16)$$

We also have the following properties.

#### Property 1

$\tilde{W}_1(K_1-1, 0) = 0$  and  $\tilde{W}_1(K_1-1, \lambda_1)$  is a monotonic increasing function with respect to  $\lambda_1$  given  $K_1-1$ , where  $K_1-1 > 0$ .

#### Property 2

$\tilde{W}_2(K-MK_1, 0) = 0$  and  $\tilde{W}_2(K-MK_1, \lambda_2)$  is a monotonic increasing function with respect to  $\lambda_2$  given  $K-MK_1$ , where  $K-MK_1 > 0$ .

Therefore,  $B(K_1-1)$  is well defined in  $C(K_1)$  and for any point  $(\lambda_1, \lambda_2) \in B(K_1-1)$ , we have

$$\begin{aligned} \tilde{W}_2(K-MK_1, \lambda_2+\Delta) > \tilde{W}_2(K-MK_1, \lambda_2), \\ \text{for any } \Delta > 0, \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} \tilde{W}_2(K-MK_1, \lambda_2 - \Delta) &< \tilde{W}_2(K-MK_1, \lambda_2) , \\ \text{for any } 0 < \Delta \leq \lambda_2 . \end{aligned} \quad (3.18)$$

The same conclusions apply to the Boundary  $B(K_1)$  and therefore  $B(K_1)$  is well defined in  $C(K_1)$  and for any point  $(\lambda_1, \lambda_2) \in B(K_1)$ , we have

$$\begin{aligned} \tilde{W}_2(K-M(K_1+1), \lambda_2 + \Delta) &> \tilde{W}_2(K-M(K_1+1), \lambda_2) , \\ \text{for any } \Delta > 0 , \end{aligned} \quad (3.19)$$

and

$$\begin{aligned} \tilde{W}_2(K-M(K_1+1), \lambda_2 - \Delta) &< \tilde{W}_2(K-M(K_1+1), \lambda_2) , \\ \text{for any } 0 < \Delta \leq \lambda_2 . \end{aligned} \quad (3.20)$$

Consider any point  $(\lambda_1, \lambda_2) \in B(K_1-1)$ . Since

$$\tilde{W}_1(K_1-1, \lambda_1) > \tilde{W}_1(K_1, \lambda_1) , \quad \lambda_1 \neq 0 , \quad (3.21)$$

We have, by the definition of  $B(K_1-1)$ ,

$$\begin{aligned} W(K_1, K-MK_1, \lambda_1, \lambda_2) \\ &= \text{Max}(\tilde{W}_1(K_1, \lambda_1), \tilde{W}_2(K-MK_1, \lambda_2)) \\ &= \tilde{W}_2(K-MK_1, \lambda_2) . \end{aligned} \quad (3.22)$$

Similarly, since

$$\tilde{W}_2(K-M(K_1-1), \lambda_2) < \tilde{W}_2(K-MK_1, \lambda_2) , \quad \lambda_2 \neq 0 , \quad (3.23)$$

We have

$$\begin{aligned} W(K_1-1, K-M(K_1-1), \lambda_1, \lambda_2) \\ &= \text{Max}(\tilde{W}_1(K_1-1, \lambda_1), \tilde{W}_2(K-M(K_1-1), \lambda_2)) \end{aligned}$$

$$= \tilde{W}_1(K_1-1, \lambda_1) . \quad (3.24)$$

By similar argument, for any point  $(\lambda_1, \lambda_2) \in B(K_1)$ , we have

$$\begin{aligned} & W(K_1+1, K-M(K_1+1), \lambda_1, \lambda_2) \\ &= \tilde{W}_2(K-M(K_1+1), \lambda_2), \end{aligned} \quad (3.25)$$

and

$$\begin{aligned} & W(K_1, K-MK_1, \lambda_1, \lambda_2) \\ &= \tilde{W}_1(K_1, \lambda_1). \end{aligned} \quad (3.26)$$

Consider any point  $(\lambda_1, \lambda_2) \in B(K_1-1)$ . Let  $\Delta_1 > 0$ . From (3.22) and (3.17), we have

$$\begin{aligned} & W(K_1, K-MK_1, \lambda_1, \lambda_2+\Delta_1) \\ &= \tilde{W}_2(K-MK_1, \lambda_2+\Delta_1). \end{aligned} \quad (3.27)$$

Incorporating (3.17) and the definition of  $B(K_1-1)$ , we have

$$\tilde{W}_2(K-MK_1, \lambda_2+\Delta_1) > \tilde{W}_1(K_1-1, \lambda_1) . \quad (3.28)$$

Since

$$\tilde{W}_2(K-MK_1, \lambda_2+\Delta_1) > \tilde{W}_2(K-M(K_1-1), \lambda_2+\Delta_1) , \quad (3.29)$$

We conclude that

$$\begin{aligned} & W(K_1, K-MK_1, \lambda_1, \lambda_2+\Delta_1) \\ &> W(K_1-1, K-M(K_1-1), \lambda_1, \lambda_2+\Delta_1) . \end{aligned} \quad (3.30)$$

Therefore,  $K_1^* \neq K_1$  at  $(\lambda_1, \lambda_2+\Delta_1)$ . Consequently, in the region bounded by  $(0, \lambda_2)$  and  $B(K_1-1)$ , we have  $K_1^* \neq K_1$ .

Consider now any point  $(\lambda_1, \lambda_2) \in B(K_1 - 1)$  and let  $0 < \Delta_2 \leq \lambda_2$ .  
From the definition of  $B(K_1 - 1)$  and by applying (3.18), we have

$$\begin{aligned} & \tilde{W}_1(K_1 - 1, \lambda_1) \\ & > \tilde{W}_2(K - MK_1, \lambda_2 - \Delta_2) . \end{aligned} \quad (3.31)$$

We also have

$$\begin{aligned} \tilde{W}_1(K_1 - i, \lambda_1) & > \tilde{W}_1(K_1 - 1, \lambda_1) > \tilde{W}_1(K_1, \lambda_1) , \\ 1 & < i < K_1 . \end{aligned} \quad (3.32)$$

Therefore,

$$\begin{aligned} & W(K_1 - i, K - M(K_1 - i), \lambda_1, \lambda_2 - \Delta_2) \\ & > W(K_1, K - MK_1, \lambda_1, \lambda_2 - \Delta_2) . \end{aligned} \quad (3.33)$$

Therefore, we conclude that in the subregion  $C(K_1)$  which is bounded by  $B(K_1 - 1)$  and  $(\lambda_1, 0)$ ,  $K_1^* \neq K_1 - i$ ,  $i \geq 1$ .

Now consider any point  $(\lambda_1, \lambda_2) \in B(K_1)$  and let  $\Delta_3 > 0$ . From the definition of  $B(K_1)$  and by applying (3.19), we have

$$\tilde{W}_1(K_1, \lambda_1) < \tilde{W}_2(K - M(K_1 + 1), \lambda_2 + \Delta_3) . \quad (3.34)$$

We also have

$$\begin{aligned} & \tilde{W}_2(K - MK_1, \lambda_2 + \Delta_3) \\ & < \tilde{W}_2(K - M(K_1 + 1), \lambda_2 + \Delta_3) . \end{aligned} \quad (3.35)$$

Futhermore,

$$\begin{aligned} & \tilde{W}_2(K - M(K_1 + 1), \lambda_2 + \Delta_3) \\ & < \tilde{W}_2(K - M(K_1 + 1), \lambda_2 + \Delta_3) , \\ 1 & < i \leq [K/M] - K_1 . \end{aligned} \quad (3.36)$$



Hence,

$$\begin{aligned}
 & W(K_1, K - MK_1, \lambda_1, \lambda_2 + \Delta_3) \\
 & < W(K_1 + 1, K - M(K_1 + 1), \lambda_1, \lambda_2 + \Delta_3) , \\
 & 1 \leq i \leq [K/M] - K_1 .
 \end{aligned} \tag{3.37}$$

Therefore, in the subregion of  $C(K_1)$  which is further bounded by  $B(K_1)$  and  $(0, \lambda_2)$ ,  $K_1^* \neq K_1 + i$ ,  $1 \leq i \leq [K/M] - K_1$ .

Consider any point  $(\lambda_1, \lambda_2) \in B(K_1)$  and let  $0 < \Delta_4 \leq \lambda_2$ . From (3.26) and (3.20), we have

$$\begin{aligned}
 & W(K_1, K - MK_1, \lambda_1, \lambda_2 - \Delta_4) \\
 & = \tilde{W}_1(K_1, \lambda_1) .
 \end{aligned} \tag{3.38}$$

By the definition of  $B(K_1)$  and by applying (3.20),

$$\begin{aligned}
 & \tilde{W}_1(K_1, \lambda_1) \\
 & > \tilde{W}_2(K - M(K_1 + 1), \lambda_2 - \Delta_4) .
 \end{aligned} \tag{3.39}$$

We also have

$$\tilde{W}_1(K_1, \lambda_1) > \tilde{W}_1(K_1 + 1, \lambda_1) . \tag{3.40}$$

Therefore,

$$\begin{aligned}
 & W(K_1, K - MK_1, \lambda_1, \lambda_2 - \Delta_4) \\
 & > W(K_1 + 1, K - M(K_1 + 1), \lambda_1, \lambda_2 - \Delta_4) .
 \end{aligned} \tag{3.41}$$

Hence, in the subregion of  $C(K_1)$  which is further bounded by  $B(K_1)$  and  $(\lambda_1, 0)$ ,  $K_1^* \neq K_1$ .

Since  $\tilde{W}_1(K_1, \lambda_1) < \tilde{W}_1(K_1 - 1, \lambda_1)$ ,  $B(K_1)$  and  $B(K_1 - 1)$  have no common point except  $(0, 0)$ . Therefore,  $\Psi(K_1) = A(K_1)$ .

Q.E.D.

When the system is implemented by a Demand-Assignment TDMA scheme, to optimally allocate the slots given  $K$  and  $M$ , we proceed as follows. Construct  $\Omega$  according to its definition. Then obtain the boundaries  $B(i)$  (defined in Theorem 3.1),  $0 \leq i \leq [K/M]$ , in  $\Omega$ . For any given pair of traffic rates  $\lambda_1$  and  $\lambda_2$ , if  $(\lambda_1, \lambda_2) \in \Omega$ , and is between  $B(i-1)$  and  $B(i)$ , set  $K_1^* = i$ ,  $1 \leq i \leq [K/M]$ .

### Numerical Results

To demonstrate the optimal allocation of slots when the system is implemented by a Demand-Assignment TDMA scheme, we consider the following numerical example.

The channel is time-slotted and there are 18 slots in every frame,  $K = 18$ . We also set  $M = 5$ ,  $\mu_1 = 0.1$  and  $\mu_2 = 1.0$ .

Since  $K = 18$  and  $M = 5$ , possible pairs of  $K_1$  and  $K_2$  are  $(1, 13)$ ,  $(2, 8)$ ,  $(3, 3)$ .

By the definition of  $\Omega(K_1)$ ,

$$\Omega(1) = \{(\lambda_1, \lambda_2): 0 \leq \lambda_1 < 0.1, 0 \leq \lambda_2 < 13\}$$

is contained in the tallest rectangle in Figure 3.1.

$$\Omega(2) = \{(\lambda_1, \lambda_2): 0 \leq \lambda_1 < 0.2, 0 \leq \lambda_2 < 8\}$$

is contained in the second tallest rectangle in Figure 3.1.

$$\Omega(3) = \{(\lambda_1, \lambda_2): 0 \leq \lambda_1 < 0.3, 0 \leq \lambda_2 < 3\}$$

is contained in the shortest rectangle in Figure 3.1.

By definition,

$$\Omega = \bigcup_{i=1}^3 \Omega(i)$$

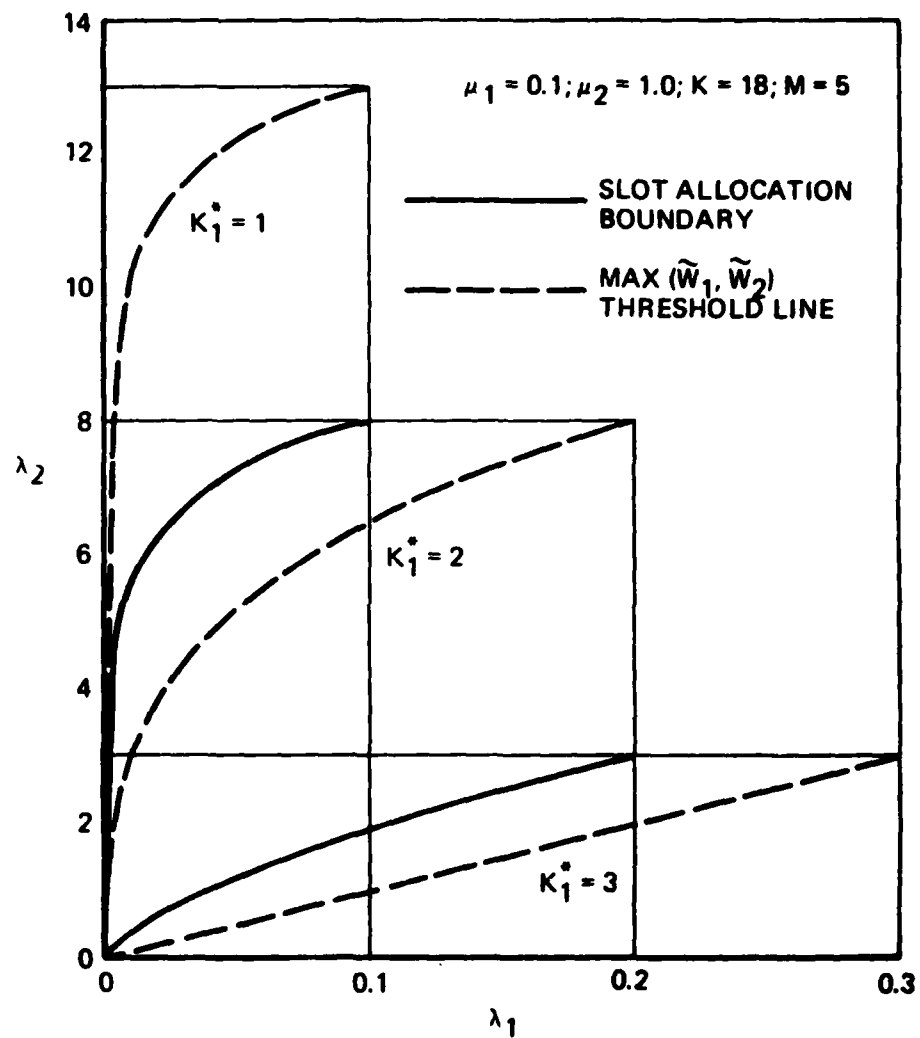


Figure 3.1. Slot Allocation under DA/TDMA Schemes with Fixed Boundaries.

and is therefore the union of the regions covered by the three rectangles in Figure 3.1.

Next, we construct the boundaries defined in Theorem 3.1.  $B(0)$  is from 0 to 13 along the ordinate.  $B(1)$  is determined by the line satisfying the equation

$$\begin{aligned}\tilde{W}_1(1, \lambda_1) &= \tilde{W}_2(8, \lambda_2), \\ 0 \leq \lambda_1 &< 0.1, \quad 0 \leq \lambda_2 < 8.\end{aligned}\tag{3.42}$$

$B(2)$  is determined by the equation satisfying the equation

$$\begin{aligned}\tilde{W}_1(2, \lambda_1) &= \tilde{W}_2(3, \lambda_2), \\ 0 \leq \lambda_1 &< 0.2, \quad 0 \leq \lambda_2 < 3.\end{aligned}\tag{3.43}$$

$B(3)$  is from 0 to 0.3 along the abscissa.

Following Theorem 3.1,  $\Psi(1)$  is the highest region in which  $K_1^* = 1$ ,  $\Psi(2)$  is the middle region in which  $K_1^* = 2$  and  $\Psi(3)$  is the lowest region in which  $K_1^* = 3$ .

Each region can be further divided into two portions by the threshold line satisfying the equations

$$\begin{aligned}\tilde{W}_1(K_1^*, \lambda_1) &= \tilde{W}_2(K_1 - MK_1^*, \lambda_2), \\ \text{for } K_1^* &= 1, 2, 3.\end{aligned}\tag{3.44}$$

From the constructions of these threshold lines, in the portion above a threshold line

$$W(K_1^*) = \tilde{W}_2(K - MK_1^*, \lambda_2),\tag{3.45}$$

and in the portion below a threshold line

$$W(K_1^*) = \tilde{W}_1(K_1^*, \lambda_1). \quad (3.46)$$

Figure 3.2 plots  $W(\cdot)$  versus  $\lambda_2$  when  $\lambda_1 = 0.14$ .  $K_1$  has to be set at least equal to 2, otherwise,

$$W(\cdot) = \tilde{W}_1(1, 0.14) = \infty.$$

From the figure,  $K_1^* = 3$  when  $0 \leq \lambda_2 \leq 2.33$ . When  $2.33 < \lambda_2$ ,  $K_1^*$  has to be switched to 2 in order to yield minimum  $W(\cdot)$ .

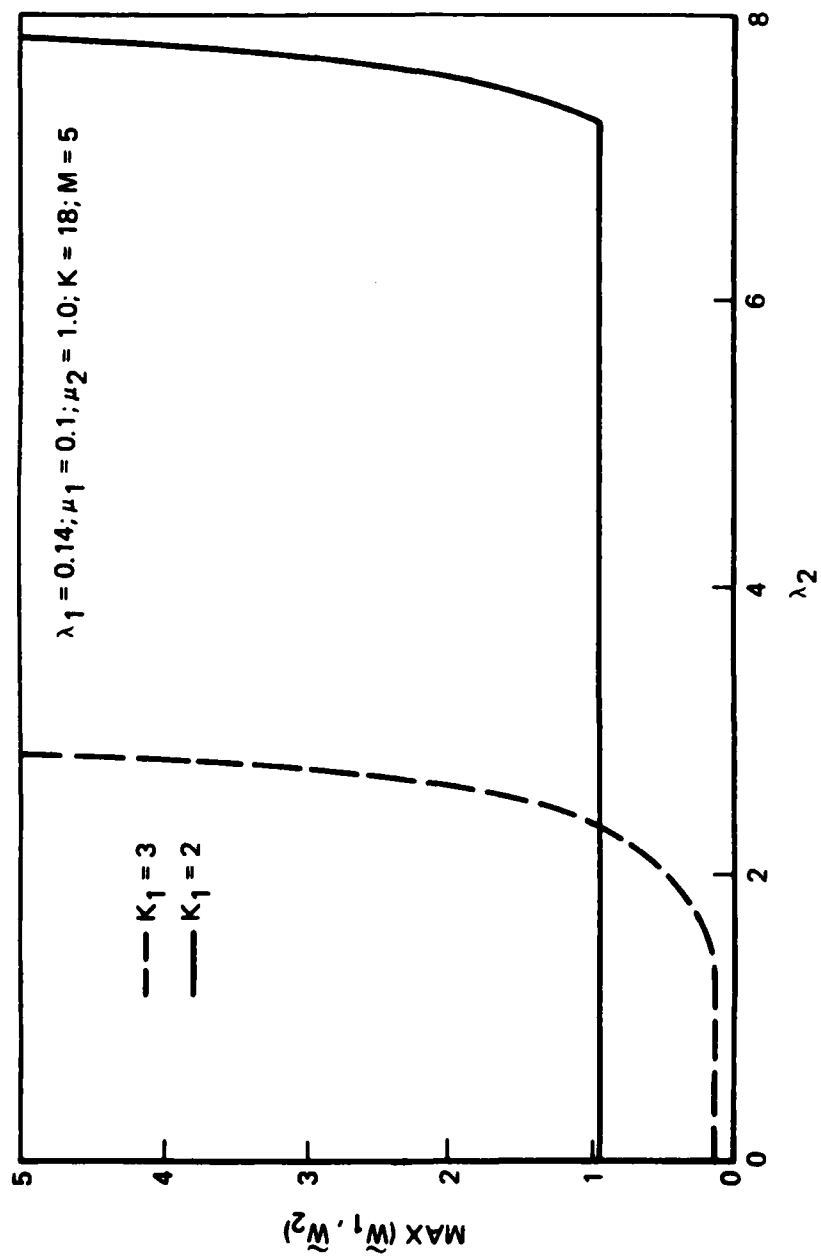


Figure 3.2. Max ( $\tilde{W}_1, \tilde{W}_2$ ) vs  $\lambda_2$  Curves under DA/TDMA Schemes with Fixed Boundaries.

## CHAPTER IV

### A CLASS OF HYBRID TDMA/COLLISION-RESOLVING SCHEMES

#### 4.1 Introduction

In this chapter, we consider the sharing of a slotted broadcast communication channel by a number of geographically distributed sources. The traffic accommodated by the channel is composed of single packet messages (if the system is implemented by a Fixed-Assignment TDMA scheme, this constraint is relaxed and the traffic is considered to be messages composed of multiple packets) arriving at the sources. Every source has the capability of listening to and receiving the packets transmitted in previous slots by all the sources, including itself. Hence, we are considering a broadcast (fully connected) packet-switched network.

Three events can happen at a channel slot. First, no source transmits any packet in a slot, then it is an empty slot. Second, a single source transmits a packet in a slot, then the packet can be received successfully by any source. This is recognized as a successful transmission. Third, two or more sources transmit packets in a single slot, then a collision occurs and none of the packets involved can be received successfully by any source. When this occurs, all the collided packets have to be retransmitted in future slots until they can be successfully received.

A source can gather information about the states of the other sources by only listening to the broadcasts from the channel. Based on this information, a scheme can be devised to allocate the channel

slots among the sources and to resolve collisions when they occur. The scheme is required to be devised also in such a way so that certain performance criteria are satisfied.

### Motivation

Three basic schemes in accessing such a broadcast communication channel are a Fixed-Assignment TDMA scheme, a Random Access scheme and a Tree Search scheme.

Under a Fixed-Assignment TDMA scheme, the channel slots are arranged to form frames of fixed durations. Each source is then given a fixed portion of every frame and contention among sources is avoided.

Under a Random Access scheme, when a packet arrives at a source, the source transmits the packet without waiting and determines whether or not the packet is involved in a collision by listening. If a collision occurred, then the packet will be retransmitted in a randomly chosen future slot. Retransmission is allocated to a randomly chosen future slot so as to avoid sure collision. This is repeated if necessary until the packet is successfully transmitted.

Under a Tree Search scheme, the channel slots are arranged to form pairs of slots called steps. The entire population of sources is divided into two groups. At the first step, the sources in the first group are given transmission rights in the first slot and those in the second group are given transmission rights in the second slot. Collision and subsequent collision(s) in each group are resolved in a binary manner. (This has been briefly discussed in Chapter I and will be



described in more detail later in this chapter.)

The performance of a Fixed-Assignment TDMA scheme is good when the throughput,  $S$ , of the channel is high, and poor when  $S$  is low. A Random Access scheme and a Tree Search scheme individually offer good performance for low values of  $S$ , but are inefficient for high values of  $S$ . Our goal is to extend these schemes to a class of hybrid access-control schemes so that optimum performance of the channel can be obtained by choosing appropriate members of the class for the entire range of  $S$ .

We observe that under a Fixed-Assignment TDMA scheme, every slot is assigned to only one source. Under a Random Access scheme or a Tree Search scheme, a slot or a pair of slots is first assigned to all the sources and then if collision occurs, an established policy is used to resolve the collision. Hence, we combine these three schemes in the following manner. We divide the entire population into groups of an equal number of sources. Then, we arrange the channel slots to form frames of fixed durations and each group is assigned a fixed portion of every frame. Hence, the groups can now be served independently in a cyclic manner. Collisions among sources in each group can be resolved by applying a Random Access scheme or a Tree Search scheme. We define the set of schemes constructed by this method a class of hybrid access-control schemes which include Fixed-Assignment TDMA schemes, Random Access schemes and Tree Search schemes. If we apply a Random Access scheme to resolve the collisions among the sources of each group and if the population is infinite, then we have a Group Random Access scheme, which has been studied extensively by

Rubin [10]. In this chapter, we apply a Tree Search scheme and the population is assumed to be finite. However, the results obtained can be applied to the case when the population is infinite.

In the next section, we will present the system models to be studied in this chapter. In Section 4.3, we will describe two policies (required by the said above class of hybrid access-control schemes) chosen to resolve collisions among the sources of each group. The intuitive reasons for choosing these policies, including the effect of propagation delay, will be discussed. In Section 4.4, we will analyze the procedures for resolving collisions by following these two policies. Numerical results will be presented. In Section 4.5, we will study the systems implemented by these two policies when the sources are assumed to have buffers of limited capacities. Numerical results will be used to compare these two policies. In Section 4.6, we will study the systems implemented by the two policies when the sources are assumed to have buffers of unlimited capacities. Exact analysis can be shown to be impossible and an approximation will be used to analyze the system. Numerical results will be used to compare these two policies. The approximation will be justified.

#### 4.2 The System Models

The following is a description of the channel model which will be used in this chapter. The class of hybrid access-control schemes used to allocate the capacity of the channel among the sources is presented. The source models which will be used in this chapter are briefly explained. Performance measures of the channel are also defined.

### The Channel Model

The channel is considered to be time-slotted, that is, the channel is divided into equal durations called slots. These slots are long enough so that exactly one packet can be transmitted in a given slot. Problems of modulation, synchronization, coding and the like are assumed to have been solved. The channel is to be shared by a finite number of sources,  $N_T$ . These sources are assumed to be synchronized to slot boundaries, therefore the channel can be considered merely as a succession of fixed time slots shared by the sources.

All the sources are considered to be alike. Although they are geographically distributed, each source can listen to the transmission of packets from all the sources, including itself. No source can obtain full information about the states of the other sources. However, a source can obtain partial information about the status of the other sources by listening to the previous slots. A slot is considered to be empty if no packet is transmitted in it by any source. If only one source transmits a packet in a slot, the transmission is regarded as a successful transmission. If two or more sources transmit packets in the same slot, a collision occurs and none of the packets can be correctly received. In the latter case, it is necessary then to arrange for these sources to retransmit the collided packets. No source can transmit more than one packet in the same slot, as a collision would be sure to occur. Every source has the ability to distinguish if a previously received slot was an empty slot or whether it was a successful transmission or a collision. Hence, every source listens to the previous slots; based on this information and a previously

agreed policy, decides whether or not to transmit in a given slot.

#### The Class of Hybrid Access-Control Schemes

Because of the possibility of having collisions of packets from different sources, it is necessary to design a scheme to assign transmission rights in every slot among the sources so as to avoid collisions and when collisions cannot be avoided, to arrange retransmission of the collided packets in future slots. The scheme is also required to satisfy certain system performance criteria such as channel throughput, the average delay of the accepted packets and the loss probability of a new arriving packet. Some of these access-control schemes have been discussed in the previous section, and were noted to yield good performance only when the throughput,  $S$ , of the channel is in the neighborhood of either 0 or 1. In the following, we will introduce a class of hybrid access-control schemes which includes the discussed schemes. Then by varying control parameters, we will study the performance of the system through the entire throughput range. Furthermore, this will expose the possibility of finding an optimum access-control scheme.

The entire population of sources,  $N_T$ , is divided into  $N_G$  groups of an equal number of sources,  $G_S$ . Hence,  $N_T = N_G G_S$ . The time slots of the channel are arranged to form consecutive blocks called frames. Each frame is composed of  $\alpha$  consecutive slots. Each slot in a frame is associated with a group of sources by its position in the frame. Hence,  $\alpha = N_G$ . For example, if the first slot is assigned to a group, then the first slot of every frame is assigned to that group. This is illustrated by Figure 4.1. This arrangement is similar to a

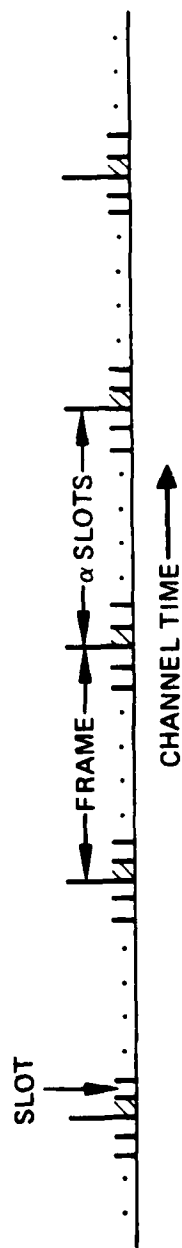


Figure 4.1. Time Frame Structure for a Class of Hybrid Access-Control Schemes.

Fixed-Assignment TDMA scheme. However, it differs from a Fixed-Assignment TDMA scheme in that each slot in a frame is associated with a group of sources. Of course, if  $N_G = N_T$ , the structure of the time frame is identical to that of a Fixed-Assignment TDMA scheme.

Now, each group is associated with a slot in every frame. To complete the design, a policy must be chosen to assign transmission rights among the sources in each group. We must determine which sources in a group should have transmission rights in a given slot assigned to the group. The details of choosing a policy may vary according to the particular environment considered. The main concept is that, once  $N_G$  is fixed and a policy is chosen, we have established an access-control scheme. Then, by varying  $N_G$ , consequently  $G_S$ , we subsequently develop a family of access-control schemes associated with that policy. Finally, by choosing different policies, we further develop families of access-control schemes which form a class of access-control schemes. In Section 4.3, we will introduce two policies developed from two versions of a Tree Search scheme. The intuitive reasons for choosing these policies will also be discussed.

#### The Source Models

Single packets are assumed to be arriving at the sources which share the communication channel. By the construction of the class of hybrid access-control schemes, buffers are required to be established at each source to store packets arriving at the source. The number of buffers required to be established at each source will be discussed in the next section. We will consider two buffer models in this chapter. In one model, the buffers are assumed to have limited

capacities and in the other model, unlimited capacities. We will study the first case in Section 4.5 and the second case in Section 4.6.

### Channel Performance Measures

Each of the systems to be considered is composed of a communication channel and the sources that share it. In order to quantitatively assess the performance of each system, several performance measures are defined. The measures which will be relevant to our study are:

$\lambda$  - the average input rate of the channel. Let  $\hat{A}_n$  be the total number of packets arriving at the sources in the  $n^{\text{th}}$  slot, then

$$\lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E[\hat{A}_n], \text{ if it exists.}$$

$S$  - the throughput of the channel. Let  $\hat{N}_n$  be the number of successful transmission in the  $n^{\text{th}}$  slot, then

$$S = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E[\hat{N}_n], \text{ if it exists.}$$

$\ell$  - the loss probability of a new arriving packet. Let  $\hat{L}_n$  be the number of packets lost in the  $n^{\text{th}}$  slot, then

$$\ell = \lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{E[\hat{L}_n]}{E[\hat{A}_n]}, \text{ if it exists.}$$

$E[\hat{Q}]$  - the average number of packets at a source (excluding the packet being transmitted). Let  $\hat{Q}_n$  be the number of packets at a source (excluding the packet being transmitted) in the  $n^{\text{th}}$  slot, then

$$E[\hat{Q}] = \lim_{n \rightarrow \infty} E[\hat{Q}_n], \text{ if it exists.}$$

$E[D]$  - the average delay of the accepted packets. Let  $D_n$  be the delay of the  $n^{\text{th}}$  accepted packet, then

$$E[D] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E[D_n], \text{ if it exists.}$$

Stability - If  $E[D]$  is finite, then the system is stable. Otherwise, it is unstable.

#### 4.3 Two Versions of a Tree Search Scheme

In Section 4.2, on the class of hybrid access-control schemes, we stated that in order to complete the design of a hybrid access-control scheme, a policy must be chosen to resolve unavoidable collisions among the sources of each group. For this purpose, we present two policies which are two versions developed from a Tree Search scheme. The first policy is motivated by its simple form and the second policy is developed from the first.

##### Policy 1

To start with, we assume there is no propagation delay, that is, every source can determine the outcome of the present slot before the commencement of the next slot. Hence, transmission rights in the next slot can be determined immediately following the start time of the present slot. Recall that the total population  $N_T$  is divided into  $N_G$  groups, each containing  $G_S$  sources. Let  $L$  be a non-negative integer. The policy requires the population to be divided in such a way that there are  $2^L$  sources in each group, that is,  $G_S = 2^L$  and  $N_T = N_G G_S$ . Since all the groups are alike and independent, we can, by symmetry, consider only one group of sources and the slots assigned



to this group in the following analysis. The process of determining the sources which are given transmission rights in a slot can be visualized by a tree diagram. (This is illustrated by an example in Figure 4.2 with  $L = 4$ .) The end nodes in the tree represent the  $G_S = 2^L$  sources in the group. For each slot, a node is selected from the tree according to certain rules to be given below. Then, all the sources (end nodes) which are connected to the selected node by sub-branches are given transmission rights in the slot. If the selected node is an end node of the tree, then only the source which is represented by this end node is given the right to transmit in the slot.

We now describe how a node is selected from the tree at each slot.

- 1) If a collision occurs in the present slot, the next node is the one at the immediate lower right of the present node.
- 2) If no collision occurs in the present slot and the present node has an immediate upper left node, then a) if the present slot is an empty slot, the next node is the immediate lower right node of the node which is at the left of the present node; b) if the present slot is a successful transmission, the next node is the one at the left of the present node.
- 3) If no collision occurs in the present slot and the present node has an immediate upper right node, we trace up in the upper right direction until a) we reach the  $L^{\text{th}}$  level node, then the next node is the  $L^{\text{th}}$  level node; b) we reach a node without an immediate upper right node, then the node to the left of this node is the next node.



4) If no collision occurs in the present slot and the present node is the  $L^{\text{th}}$  level node, the next node is the  $L^{\text{th}}$  level node.

Note that in rule 2), if the present slot is an empty slot, following b) will yield a sure collision, therefore a) is followed.

This policy is easy to implement. It operates in a cyclic manner and always begins at the  $L^{\text{th}}$  level node, that is, all sources are given transmission rights. If no collision occurs, the next node is the  $L^{\text{th}}$  level node again. If a collision occurs, sub-nodes will be selected, splitting the sources in a binary manner until said collision and other (if any) subsequent collision(s) are resolved.

We now discuss why the splittings are done in a binary manner. We assume the total traffic of the group of sources to be low enough such that if a collision occurs, it is more likely to be caused by the simultaneous transmission of two packets rather than being caused by more than two packets. If a collision occurs at the  $L^{\text{th}}$  level node and  $k$  of the  $2^L$  sources are given transmission rights in the next slot, the probability of a successful transmission in the next slot is given by an element of the Hypergeometric distribution

$$\frac{\binom{k}{1} \binom{2^L - k}{1}}{\binom{2^L}{2}}.$$

This expression assumes its maximum when  $k$  is equal to  $2^{L-1}$ . Also, when  $k$  is equal to  $2^{L-1}$ , the average number of packets transmitted in the next slot is equal to unity. Hence, we design the policy to split the sources in a binary manner. This requirement on the total traffic of the group of sources can be met to a certain degree by varying  $G_s$ ,

consequently  $N_G$ . If the total traffic is too high to allow slots wasted on collisions, we adjust  $N_G$  to be equal to  $N_T$ , and a Fixed-Assignment TDMA scheme is procured.

#### Propagation Delay and the Construction of Superframes

So far, we have assumed no propagation delay. The importance of propagation delay is seen in the implementation of Policy 1 when the selection of the next node depends on the outcome of the present slot. Recall that every group of sources is assigned one slot in every frame. Set  $R$  slots to be the propagation delay of the channel. If  $R < \alpha$ , then there is no problem. Otherwise, the required information about the present slot is not yet available when selecting the next node. In that case, to make the policy feasible, we construct superframes. Suppose  $n$  is the smallest integer such that  $R < n\alpha$ , then the frames are arranged to form consecutive superframes, each composed of  $n$  consecutive frames. Set  $\beta$  slots to be the duration of a superframe, then  $\beta = n\alpha$ . (This is illustrated by an example in Figure 4.3.) Hence, every group of sources is assigned  $n$  evenly spread slots in each superframe. Again, by symmetry, we consider only one group of sources and the slots assigned to this group. Due to propagation delay, the soonest slot in which the transmission rights can be determined by the outcome of a given slot of the group is the slot with the same position of the given slot in the next superframe. To satisfy this propagation constraint, we do the following. There are  $n$  slots assigned to the group in every superframe, therefore, we equip each source with  $n$  independent buffers, one buffer associated with each slot in the superframe. When a source is given transmission right in a slot, it can

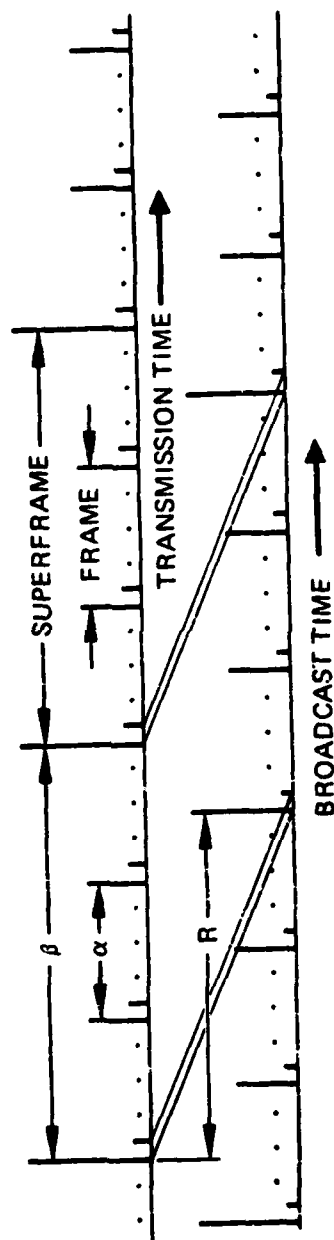


Figure 4.3. An Example of Superframe Construction. Propagation Delay is  $R$  Slots and  $2\alpha < R < 3\alpha$ .

transmit only the packet, if any, stored in the buffer associated to that slot. Also, every buffer accommodates only the packets arriving in the  $\alpha$  slots (the duration of one frame) preceding its associated slot. At this point, it is appropriate to make the following definition.

#### A Step

Given a Tree Search scheme and its superframe structure, a buffer's associated slot(s) in a superframe is defined to be a step of the buffer.

By definition, the packets stored in a buffer can only be transmitted in the steps of the buffer. Furthermore, the intervals between the starts of the successive steps of a buffer are each  $\beta$  slots (the duration of one superframe) long. We are now ready to investigate the delay of an accepted packet.

#### The Delay of an Accepted Packet

In Section 4.1, we defined some performance measures. One of them is the average delay of the accepted packets. Let's now observe not all the accepted packets, but just one, and see what its delay is made up of. We define the delay of an accepted packet to be from the start of the slot right after its arrival to the time it is successfully received. The delay of an accepted packet can be decomposed into three parts; the waiting time of the packet, the slot in which the packet is successfully transmitted and the channel's propagation delay. The last two parts are obvious. We investigate the first part.

The waiting time of a packet is defined to be from the start of the slot right after its arrival at a source to the start of the slot in which it is successfully transmitted. The waiting time of a packet can be further decomposed into three components as follows:

#### The Frame Latency of a Packet

The first component, denoted by  $W_S^{(1)}$  slots, is the frame latency of the packet. It begins at the start of the slot right after the packet's arrival at a source until the start of the first slot which is associated to the group the source belongs.  $W_S^{(1)}$  depends on the position of the slot in which the packet arrives, and the frame duration.

#### The Passive Waiting Time of a Packet

The second component, denoted by  $W_S^{(2)}$  slots, is the passive waiting time of the packet. It begins right after the packet's  $W_S^{(1)}$  period until the start of the slot in which the packet is transmitted for the first time. Let  $W^{(2)}$  be the number of steps required to transmit the packets being in process and the packets in the source buffer which have arrived at an earlier time. Then  $W_S^{(2)} = \beta W^{(2)}$ . Hence,  $W_S^{(2)}$  is directly proportional to  $\beta$ .  $W^{(2)}$  is defined to be the passive waiting steps of the packet.

#### The Active Waiting Time of a Packet

The third component, denoted by  $W_S^{(3)}$  slots, is the active waiting time of the packet. It begins right after the packet's  $W_S^{(2)}$  period until the start of the slot in which the packet is successfully transmitted. Let  $W^{(3)}$  be the number of steps after  $W_S^{(2)}$  but before the slot in which the packet is successfully transmitted. Then  $W_S^{(3)} = \beta W^{(3)}$ .

Hence,  $W_S^{(3)}$  is also directly proportional to  $\beta$ .  $W^{(3)}$  is defined to be the active waiting steps of the packet.

Hence, we conclude that the delay of an accepted packet is given by

$$D = W_S^{(1)} + (W^{(2)} + W^{(3)})\beta + I + R. \quad (4.1)$$

Given a Tree Search scheme and its superframe structure, it is observed that the number of steps,  $W^{(2)} + W^{(3)}$ , required to resolve collisions and the superframe duration,  $\beta$ , are very important parameters when considering the delay of an accepted packet. This observation motivates the construction of Policy 2 which is described as follows.

#### Policy 2

When Policy 1 is implemented, each group of sources is associated with one slot per frame. This frame structure is modified by associating each group of sources with two consecutive slots per frame. Hence,  $\alpha = 2N_G$ . Let  $L$  be a positive integer. Policy 2 requires the population to be divided in such a way that each group contains  $2^L$  sources, that is,  $G_S = 2^L$ . The groups are independent, by symmetry, we again consider only one group of sources and the slots associated with this group. The assignment of transmission rights among a group of sources is similar to that when Policy 1 is implemented. Policy 2 can again be visualized by a tree diagram. (This is illustrated by an example in Figure 4.4 in which  $L = 4$ .) The  $2^L$  end nodes represent the sources of a group. For each pair of slots, a pair of nodes (connected by a horizontal branch) is selected from the tree according



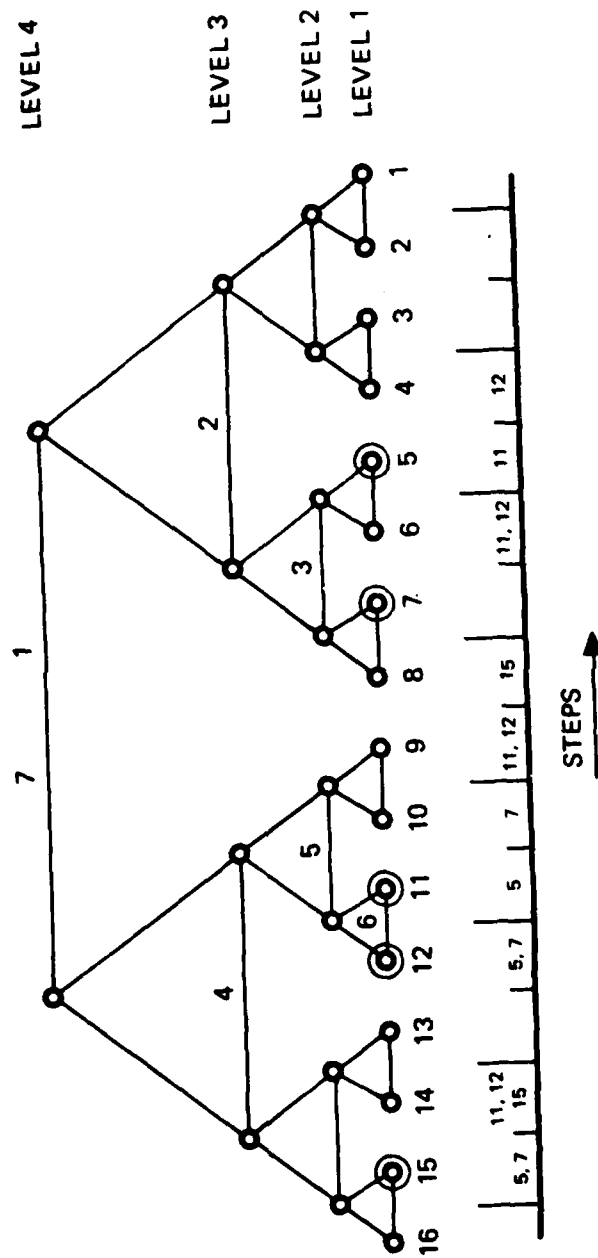


Figure 4.4. The Tree Structure of a Group of Sources under Policy 2,  $L = 4$ , with an Example illustrating the Selection of Successive Nodes.

to certain rules to be given below. Then all the sources (end nodes) which are connected by sub-branches to the right node of the selected pair are given transmission rights in the first slot and all the sources (end nodes) which are connected by sub-branches to the left node of the selected pair are given transmission rights in the second slot. If the selected pair are end nodes of the tree, then only the source represented by the right node of the pair is given transmission right in the first slot and only the source represented by the left node of the pair is given transmission right in the second slot.

We now describe how a pair of nodes is selected from the tree.

- 1) We begin at the  $L^{\text{th}}$  level pair. If no collision occurs, the next pair is again the  $L^{\text{th}}$  level pair. Otherwise, the following rules are applied until all collision(s) and subsequent collision(s), if any, are resolved. Then, we begin again at the  $L^{\text{th}}$  level pair.

- 2) For any pair of slots, resolve the collision, and subsequent collision(s), if any, in the first slot (the right node of the pair) before those in the second slot (the left node of the pair).

- 3) If the collision in a node is to be resolved, the pair of nodes to be selected next is right below the node.

Note that Policy 1 requires only to know the position of the present node and the outcome of the present slot in order to determine the position of the next node. Policy 2 requires not only to know the position of the present pair of nodes and the outcomes of the present pair of slots, it also requires the memory of the past unresolved collisions because of rule 2).

Like Policy 1, if the propagation delay of the channel is large, we need to construct superframes for Policy 2. From the structure of the frame, if  $R + 1 < \alpha$ , the construction of superframes is not necessary. Otherwise, we construct superframes as follows. Let  $n$  be the smallest integer such that  $R + 1 < n\alpha$ . Then the channel frames are arranged to form consecutive superframes, each composed of  $n$  consecutive frames, that is,  $\beta = n\alpha$ . Hence, each group of sources is associated with  $n$  evenly spread pairs of slots per superframe. Similar to Policy 1, to satisfy the propagation delay constraint, each source is equipped with  $n$  independent buffers, one buffer associated with each pair of slots in the superframe. When a source is given transmission right in one of a pair of slots, it can only transmit the packet, if any, stored in the buffer associated to that pair of slots. Also, every buffer accommodates only the packets arriving in the  $\alpha$  slots (the duration of one frame) preceding its associated pair of slots.

Note that under Policy 2, by definition, a step of a buffer is made up of two consecutive slots.

Note also that because of the similarities of the procedures in resolving collisions under both policies, the delay and the waiting time of an accepted packet under Policy 2 can be decomposed like that of the delay and the waiting time of an accepted packet under Policy 1. Hence, under Policy 2, if a packet is successfully transmitted in the first slot of a step, the packet's delay can be decomposed as

$$D = W_S^{(1)} + (W^{(2)} + W^{(3)})\beta + 1 + R \quad (4.2)$$

If a packet is successfully transmitted in the second slot of a step, then the packet's delay can be decomposed as

$$D = W_S^{(1)} + (W^{(2)} + W^{(3)})\beta + 2 + R \quad . \quad (4.3)$$

Now, let's discuss the motivation for the construction of Policy

2. We mentioned previously that if a collision occurs, it is more likely to be caused by the simultaneous transmission of two packets than more than two packets. A collision caused by the simultaneous transmission of two packets requires very likely three steps (three slots in three successive frames) to resolve when Policy 1 is implemented. If Policy 2 is implemented, very likely two steps (four slots in two successive frames) are required. The retransmission delay incurred in resolving a collision depends on the number of steps required and the duration of the intervals between successive steps. If propagation delay is small such that the construction of superframes is not necessary, then the retransmission delays incurred in resolving a collision under Policy 1 is likely to be less than that under Policy 2 because the frame duration of Policy 2 is twice as long as the frame duration of Policy 1. However, if propagation delay is large, retransmission of packets are made at intervals of superframes. If the durations of the superframes of the two policies are the same or close, then the retransmission delays incurred in resolving a collision under Policy 2 is likely to be less than that under Policy 1 because Policy 2 requires less steps than does Policy 1 to resolve a collision.

(This is illustrated in Figure 4.5.)

Obviously, Policy 2 has two disadvantages. First, it cannot

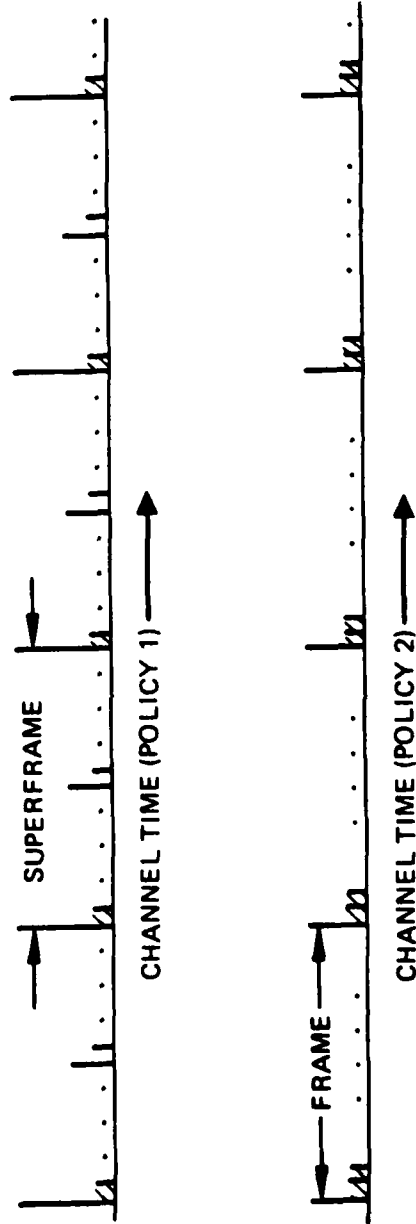


Figure 4.5. An Example illustrating the Effect of Propagation Delay under Policy 1 and Policy 2.  $N_G$  and  $R$  are of Fixed Values. Under Policy 1,  $\alpha = N_G$  and  $\alpha < R < 2\alpha$ . Under Policy 2,  $\alpha = 2N_G$  and  $R+1 < \alpha$ .

save slots which can be saved by rule 2)a) of Policy 1. Second, each group of sources is associated with two consecutive slots per frame which automatically adds an additional delay of one slot to every packet successfully transmitted in the second slot of a pair.

#### 4.4 Collision-Resolving Procedures

In Section 4.3, we presented two policies to resolve collisions incurred in each group of sources. In this section, we analyze the procedures involved in resolving collisions by each of the two policies.

According to either policy, a sequence of sources from a group to be given transmission rights in a sequence of steps is equivalent to a sequence of jumps upon the nodes (pairs of nodes) of the tree constructed from the group. Henceforth, we will call a jump upon a node (pair of nodes) a step. We make the following definitions and assumption.

##### Group Level

Given a group of  $2^L$  sources and its tree structure, if at a step, a  $k^{\text{th}}$  level node is selected, we say the group is at level  $k$ .

##### An Epoch

Given a group of  $2^L$  sources and its tree structure, an epoch is a sequence of steps starting at the  $L^{\text{th}}$  level node of Policy 1 (pair of nodes of Policy 2) and ending at the node (pair of nodes) preceding the next  $L^{\text{th}}$  level node (pair of nodes). If the next node (pair of nodes) after the  $L^{\text{th}}$  level node (pair of nodes) is again the  $L^{\text{th}}$  level node (pair of nodes), then the epoch is composed of one step, the first  $L^{\text{th}}$  level node (pair of nodes).

### Assumption

The packets processed during an epoch are those transmitted at the first step of the epoch. Hence, no packet can join an epoch after the first step of the epoch.

### A j-epoch

An epoch during which  $j$  packets are transmitted is defined to be a  $j$ -epoch.

Note that from the above assumption, at most one packet from each source can be transmitted during an epoch. Hence, given a group of  $2^L$  sources, possible values of  $j$  are  $0, 1, 2, \dots, 2^L$ .

### Analysis of a j-epoch

Given a group of  $2^L$  sources, to study the probabilistic properties of a  $j$ -epoch, we define the following:

$T_L | j$  - the number of steps of a  $j$ -epoch.

$W_L | j$  - the sum of the active waiting steps of the packets transmitted during a  $j$ -epoch.

According to either policy, if  $j > 1$ , the  $2^L$  sources are split into two smaller groups of an equal number of sources. Among the  $j$  sources which transmit packets during the epoch, suppose  $j'$  are in the first group and  $j-j'$  are in the second group. We further define the following:

$T_L | j, j'$  - the number of steps of a  $j$ -epoch given that among the  $j$  sources which transmit packets during the epoch,  $j'$  are in the first group and  $j-j'$  are in the second group.

$W_L|j,j'$  - the sum of the active waiting steps of the packets transmitted during a  $j$ -epoch given that among the  $j$  sources which transmit packets during the epoch,  $j'$  are in the first group and  $j-j'$  are in the second group.

In the following, we will study  $T_L|j$ ,  $W_L|j$ ,  $T_L|j,j'$  and  $W_L|j,j'$  under Policy 1 and Policy 2.

#### Policy 1

If  $L = 1$  and  $j < 2$ , we have

$$T_L|j = 1, \quad (4.4)$$

and  $W_L|j = 0. \quad (4.5)$

If  $L = 1$  and  $j = 2$ , we have

$$T_L|2 = 3, \quad (4.6)$$

and  $W_L|2 = 3. \quad (4.7)$

If  $L > 1$ , the policy can be described by the finite state machine in Figure 4.6. The states represent:

$L$  - the group is at level  $L$ .

$L-1, R$  - the group is at level  $L-1$ , right node.

$L-1, L$  - the group is at level  $L-1$ , left node.

$L-2$  - the group is at level  $L-2$ .

The weights of the edges represent:

0 - an empty slot.

1 - a successful transmission.

0, 1 - an empty slot or (exclusive) a successful transmission.

c - a collision.

r - the resolution of a previous collision at level  $L-2$  or



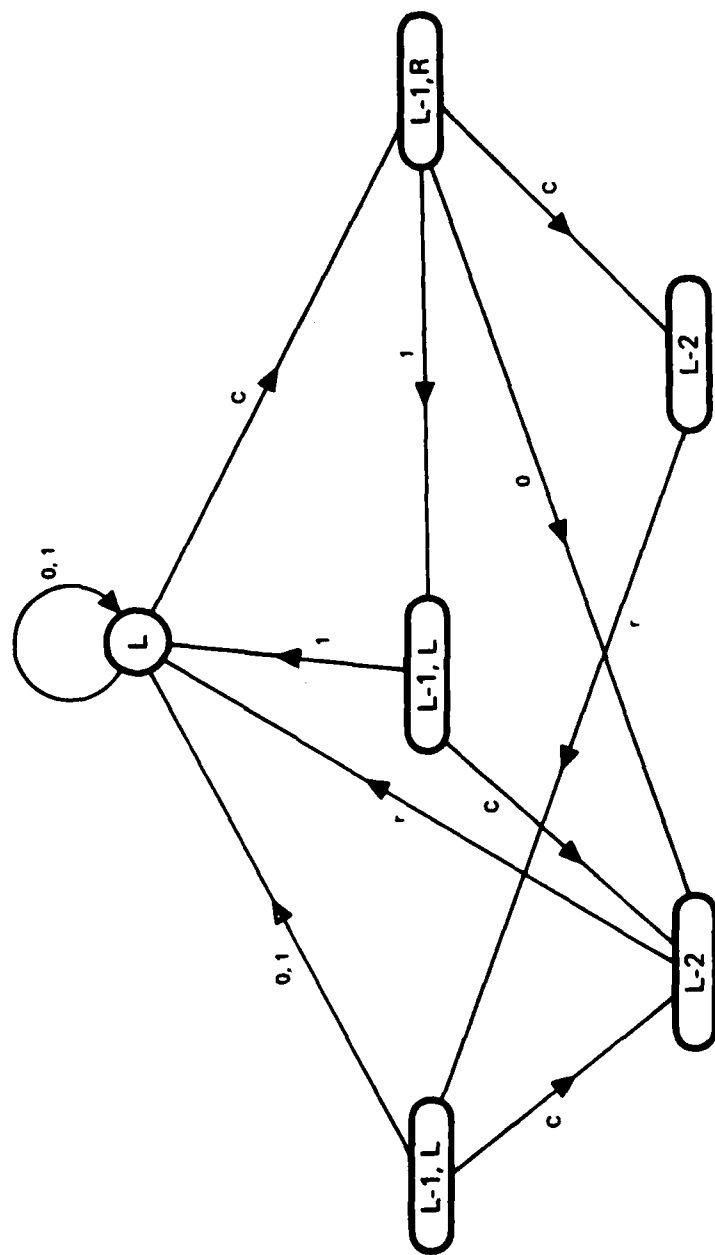


Figure 4.6. Collision-Resolving Procedures under Policy 1 when  $L > 1$ .

lower if required.

In the machine, each cycle starts from state L and goes back represents a complete epoch. The cycles represent:

For Cycle 0 or 1,  $j < 2$ . We have

$$T_L|j = 1, \quad (4.8)$$

$$\text{and } W_L|j = 0. \quad (4.9)$$

For Cycle c11,  $j = 2$  and is split into 1, 1. We have

$$T_L|2,1 = 3, \quad (4.10)$$

$$\text{and } W_L|2,1 = 3. \quad (4.11)$$

For Cycle clcr,  $j > 2$  and is split into 1,  $j-1$ . We have

$$T_L|j,1 = 1 + 1 + (T_{L-1}|j-1), \quad (4.12)$$

$$\text{and } W_L|j,1 = j + (j-1) + (W_{L-1}|j-1). \quad (4.13)$$

For Cycle c0r,  $j > 1$  and is split into 0,  $j$ . We have

$$T_L|j,0 = 1 + (T_{L-1}|j), \quad (4.14)$$

$$\text{and } W_L|j,0 = j + (W_{L-1}|j). \quad (4.15)$$

For Cycle ccr0,  $j > 1$  and is split into  $j$ , 0. We have

$$T_L|j,j = 1 + (T_{L-1}|j) + 1, \quad (4.16)$$

$$\text{and } W_L|j,j = j + (W_{L-1}|j). \quad (4.17)$$

For Cycle ccrl,  $j > 2$  and is split into  $j-1$ , 1. We have

$$T_L|j,j-1 = 1 + (T_{L-1}|j-1) + 1, \quad (4.18)$$

$$\text{and } W_L|j,j-1 = j + (W_{L-1}|j-1) + (T_{L-1}|j-1). \quad (4.19)$$

For Cycle ccrer,  $j > 3$  and is split into  $j'$ ,  $j-j'$  such that  $j' > 1$  and  $j-j' > 1$ . We have

$$T_L | j, j' = 1 + (T_{L-1} | j') + (T_{L-1} | j-j'), \quad (4.20)$$

$$\text{and } W_L | j, j' = j + (W_{L-1} | j') + (j-j')(T_{L-1} | j') \\ + (W_{L-1} | j-j'). \quad (4.21)$$

Note that  $T_{L-1} | j'$  and  $T_{L-1} | j-j'$  in Equation (4.20) are independent random variables.

### Policy 2

If  $L = 1$  and  $j < 3$ , we have

$$T_L | j = 1, \quad (4.22)$$

$$\text{and } W_L | j = 0. \quad (4.23)$$

If  $L > 1$ , the policy can be described by the finite state machine in Figure 4.7. The states represent the level of the group. The weights of the edges with brackets separated by commas indicate possible (exclusive) outcomes of a step. The first and second elements in each bracket indicate the outcomes of the first slot and the second slot, respectively, of a step.  $r$  is the resolution of a collision at level  $L-1$  or lower if required. In the concatenation of 2  $r$ s, the first  $r$  and the second  $r$  represent the resolutions of collisions in the first slot and the second slot, respectively, of a previous step at level  $L-1$  or lower if required. Each cycle starts from state  $L$  and goes back represents a complete epoch. The cycles represent:

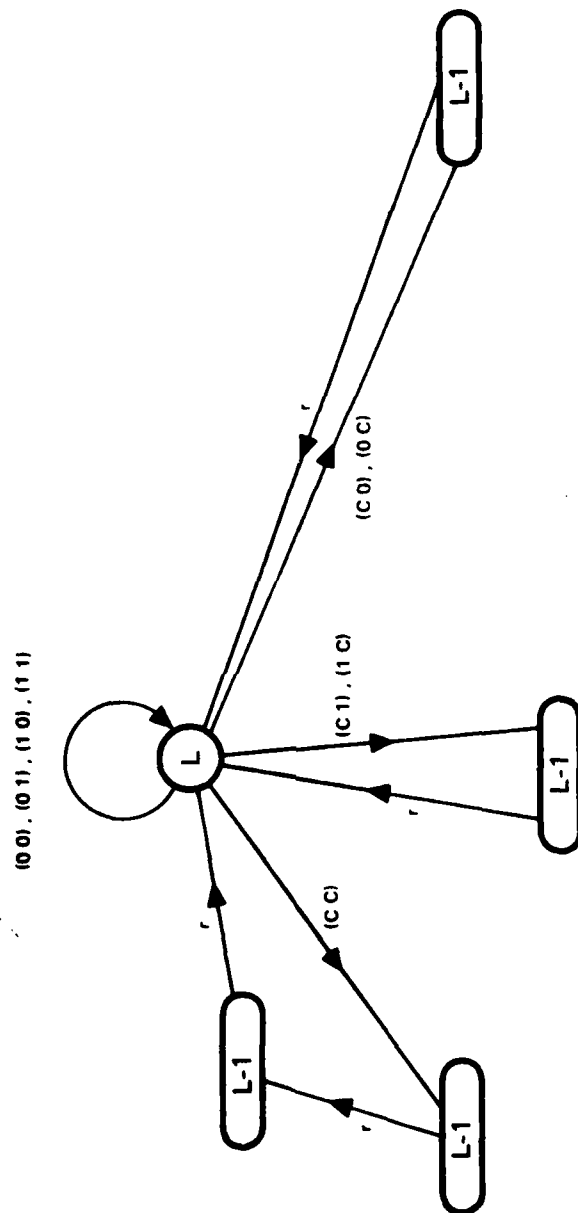


Figure 4.7. Collision-Resolving Procedures under Policy 2 when  $L > 1$ .

For Cycle (00) or (01) or (10) or (11),  $j = 3$  and is split into 0, 0 or 0, 1 or 1, 0 or 1, 1. We have

$$T_L|j = 1, \text{ if } j = 2, \quad (4.24)$$

$$\text{and } W_L|j = 0, \text{ if } j = 2. \quad (4.25)$$

$$T_L|2,1 = 1, \text{ if } j = 2, \quad (4.26)$$

$$\text{and } W_L|2,1 = 0, \text{ if } j = 2. \quad (4.27)$$

For Cycle (c0)r or (0c)r,  $j > 1$  and is split into  $j, 0$  or  $0, j$ . We have

$$T_L|j,j = T_L|j,0 = 1 + (T_{L-1}|j), \quad (4.28)$$

$$\text{and } W_L|j,j = W_L|j,0 = j + (W_{L-1}|j). \quad (4.29)$$

For Cycle (c1)r or (1c)r,  $j > 2$  and is split into  $j-1, 1$  or  $1, j-1$ . We have

$$T_L|j,1 = T_L|j,j-1 = 1 + (T_{L-1}|j-1), \quad (4.30)$$

$$\text{and } W_L|j,1 = W_L|j,j-1 = j-1 + (W_{L-1}|j-1). \quad (4.31)$$

For Cycle (cc)rr,  $j > 3$  and is split into  $j'$  and  $j-j'$  such that  $j' > 1$  and  $j-j' > 1$ . We have

$$T_L|j,j' = 1 + (T_{L-1}|j') + (T_{L-1}|j-j'), \quad (4.32)$$

$$\begin{aligned} \text{and } W_L|j,j' = j + (W_{L-1}|j') + (j-j')(T_{L-1}|j') \\ + (W_{L-1}|j-j'). \end{aligned} \quad (4.33)$$

Note that  $T_{L-1}|j'$  and  $T_{L-1}|j-j'$  in Equation (4.32) are independent random variables.

#### Theorem 4.1

Assume a system implemented by a Policy 1 or Policy 2 scheme with  $2^L$  sources in each group. Consider a  $j$ -epoch. The distribution of  $T_L|j$ , denoted by  $P_{T_L|j}(k|j)$  and the mean of  $W_L|j$ , denoted by  $E[W_L|j]$ , can be calculated numerically by recursion formulas:

$$P_{T_L|j}(k|j) = \sum_{j'=0}^j P_{T_L|j,j'}(k|j,j') \frac{\binom{2^{L-1}}{j'} \binom{2^{L-1}}{j-j'}}{\binom{2^L}{j}}, \quad (4.34)$$

and

$$E[W_L|j] = \sum_{j'=0}^j E[W_L|j,j'] \frac{\binom{2^{L-1}}{j'} \binom{2^{L-1}}{j-j'}}{\binom{2^L}{j}}, \quad (4.35)$$

where  $P_{T_L|j,j'}(k|j,j')$  and  $E[W_L|j,j']$  can be represented in terms of  $P_{T_{L-1}|j'}(k|j')$ ,  $P_{T_{L-1}|j-j'}(k|j-j')$ ,  $E[W_{L-1}|j']$  and  $E[W_{L-1}|j-j']$  according to Equations (4.4)-(4.21) when a Policy 1 scheme is implemented, and according to Equations (4.22)-(4.33) when a Policy 2 scheme is implemented.

#### Proof

Consider Policy 1.

If  $j < 2$ , Cycle 0 or Cycle 1 determines  $P_{T_L|j}(k|j)$  and  $E[W_L|j]$  according to Equations (4.4)-(4.5).

If  $j > 1$  and  $L = 1$ ,  $P_{T_L|j}(k|j)$  and  $E[W_L|j]$  are determined by Equations (4.6)-(4.7).

If  $j > 1$  and  $L > 1$ , consider the splitting of the  $2^L$  sources into two smaller groups of an equal number of sources. Among the  $j$  sources

which transmit packets during the epoch, the probability of  $j'$  are in the first group and  $j-j'$  are in the second group is given by an element of the Hypergeometric distribution

$$\frac{\binom{2^{L-1}}{j'} \binom{2^{L-1}}{j-j'}}{\binom{2^L}{j}} .$$

The splitting of  $j$  into  $j'$  and  $j-j'$  determines the cycle which represents the epoch. From the equations associated with the cycle,  $P_{T_L|j,j'}(k|j,j')$  and  $E[W_L|j,j']$  can be represented in terms of  $P_{T_{L-1}|j'}(k|j')$ ,  $P_{T_{L-1}|j-j'}(k|j-j')$ ,  $E[W_{L-1}|j']$  and  $E[W_{L-1}|j-j']$ .

The proof of the theorem for Policy 2 is similar to that for Policy 1, and is omitted here.

Q.E.D.

#### Numerical Results

Tables 4.1-4.4 show the numerical results of  $P_{T_L|j}(k|j)$  and  $E[W_L|j]$  when  $j$  is equal to 2. Notice that  $P_{T_L|j}(k|j)$  is stretched as  $L$  is increased from 1 to 6. Note that each step in Policy 1 is one slot and each step in Policy 2 is two slots.

#### 4.5 Sources with Limited Buffer Capacities

We are now ready to study the performance of the systems implemented by Policy 1 and Policy 2. As mentioned in Sections 4.1, the buffers at each source may have limited or unlimited capacities. In this section, we will study the case in which the source buffers are assumed to have limited capacities. First, we describe the source

$P_{T_L 2}(k 2)$						
$k \backslash L$	1	2	3	4	5	6
3	1.0	0.667	0.571	0.533	0.516	0.508
4	↑	0.167	0.143	0.133	0.129	0.127
5	↑	0.167	0.179	0.167	0.161	0.159
6	↑	↑	0.071	0.075	0.073	0.071
7	↑	↑	0.036	0.058	0.058	0.058
8	0.0	↑	↑	0.025	0.032	0.032
9	↑	0.0	↑	$0.83 \times 10^{-2}$	0.020	0.022
10	↑	↑	↑	↑	$0.80 \times 10^{-2}$	0.013
11	↑	↑	0.0	↑	$0.20 \times 10^{-2}$	$0.69 \times 10^{-2}$
12	↑	↑	↑	0.0	0.0	$0.25 \times 10^{-2}$
13	↑	↑	↑	↑	0.0	$0.50 \times 10^{-3}$
elsewhere	↑	↑	↑	↑	0.0	0.0

Table 4.1.  $P_{T_L|2}(k|2)$  under Policy 1.

L	1	2	3	4	5	6
$E[W_L 2]$	3.00	3.667	4.143	4.467	4.677	4.810

Table 4.2.  $E[W_L|2]$  under Policy 1.



		$P_{T_L 2}(k 2)$					
$k \backslash L$		1	2	3	4	5	6
1		1.0	0.667	0.571	0.533	0.516	0.508
2	↑		0.333	0.286	0.267	0.258	0.254
3	↑			0.143	0.133	0.129	0.127
4	↑	0.0	0.0		0.067	0.065	0.064
5	↑			0.0	0.0	0.032	0.032
6	↑				0.0	0.0	0.016
elsewhere	↓	↓	↓	↓	0.0	0.0	0.0

Table 4.3.  $P_{T_L|2}(k|2)$  under Policy 2.

L	1	2	3	4	5	6
$E[W_L 2]$	0.0	0.667	1.143	1.467	1.677	1.810

Table 4.4.  $E[W_L|2]$  under Policy 2.

buffers and the statistics of the arrivals of packets at the sources. Then, we analyze the systems implemented by Policy 1 and Policy 2.

### The Source Model

In a given system, the groups of sources are independent. Therefore, we again, by symmetry, consider only one group of sources and the slots associated to this group. Recall that every buffer accommodates only the packets arriving in the  $\alpha$  slots preceding each of its steps. Each buffer is assumed to have two buffer spaces. The first buffer space is used to store the packet being transmitted during an on-going epoch. If a source does not have any packet to transmit during an epoch, this buffer space is left empty. The second buffer space is used to store the first packet which arrives after an epoch has started. Those packets which arrive after the first are assumed to be lost. A packet in the second buffer space is transferred to the first buffer space just before the start of the next epoch.

The numbers of packets which arrive at a source in a sequence of slots are assumed to be independent and identically distributed. Let  $A$  be the number of packets arriving at a source in a slot. Set  $P_A(i)$  to be the distribution of  $A$  and  $C_A(z)$  to be the  $z$ -transform of  $P_A(i)$ , that is,

$$C_A(z) = \sum_{i=1}^{\infty} P_A(i) z^i, \quad |z| \leq 1.$$

### Analysis of the Systems

The average input rate of the channel is

$$\lambda = N_T E[A] . \quad (4.36)$$

We set  $p$  to be the probability of having at least one packet arriving in  $\alpha$  slots and  $q$  to be the probability of having no packet arriving in  $\alpha$  slots. Then

$$p = 1 - (P_A(0))^\alpha , \quad (4.37)$$

$$\text{and } q = 1 - p . \quad (4.38)$$

### Theorem 4.2

Assume a system implemented by a Policy 1 scheme or a Policy 2 scheme with  $2^L$  sources in each group. Set  $X_n$  to be the number of packets transmitted during the  $n^{\text{th}}$  epoch. Then  $\{X_n, n > 0\}$  is a Markov Chain over the state space  $\{0, 1, 2, \dots, 2^L\}$ . If  $p > 0$ , this Markov Chain is irreducible, aperiodic and positive recurrent with a state transition distribution given by

$$\begin{aligned} P(X_{n+1} = i | X_n = j) \\ = \binom{2^L}{i} p_j^i (1-p_j)^{2^L-i} , \end{aligned} \quad (4.39)$$

where

$$p_j = \sum_{k=1}^{\infty} P_{T_L|j}(k|j) (1-q^k) . \quad (4.40)$$

Furthermore, this Markov Chain has a steady state distribution  $\{\pi_j, 0 \leq j \leq 2^L\}$  which is the unique solution of the following set of linear equations:

$$\pi_j = \sum_{i=0}^{2^L} \pi_i \binom{2^L}{j} p_i^j (1-p_i)^{2^L-j}, \quad 0 \leq j \leq 2^L, \quad (4.41)$$

and

$$\sum_{j=0}^{2^L} \pi_j = 1, \quad \pi_j \geq 0. \quad (4.42)$$

### Proof

We set  $p_j$  to be the probability of having at least one arriving packet at a source in a  $j$ -epoch. The arrivals of packets at a source in different slots are independent, therefore,  $p_j$  is given by Equation (4.40). Since the arrivals of packets at sources are independent, the state transition distribution of the process  $\{X_n, n > 0\}$  is given by Equation (4.39). By Equation (4.39), we conclude that the process  $\{X_n, n > 0\}$  is Markovian. The state space of the Markov Chain  $\{X_n, n > 0\}$  is finite. If  $p > 0$ , the probability of going from every state to every other state is non-zero, therefore, the Markov Chain  $\{X_n, n > 0\}$  is irreducible, aperiodic and positive recurrent. Since the Markov Chain  $\{X_n, n > 0\}$  is irreducible, aperiodic and positive recurrent, it has a steady state distribution which is the unique solution of the set of linear Equations (4.41)-(4.42).

Q.E.D.

As described in Section 4.3, the waiting time of an accepted packet can be decomposed into three components. We now analyze these three components.

The Mean of  $W_S^{(1)}$

Consider the  $\alpha$  slots preceding a step of a buffer.

If  $\alpha = 1$ , then  $E[W_S^{(1)}] = 0$ .

If  $\alpha > 1$ , then the conditional probability of the first packet arriving in the  $i^{\text{th}}$  slot given there is at least one packet arriving in the  $\alpha$  slots is

$$\frac{(1-P_A(0))P_A(0)^{i-1}}{p}.$$

Then the mean of  $W_S^{(1)}$  is

$$\begin{aligned} E[W_S^{(1)}] &= \sum_{i=1}^{\alpha} \frac{(1-P_A(0))P_A(0)^{i-1}(\alpha-i)}{p} \\ &= \frac{\alpha}{p} - \frac{1}{1-P_A(0)}, \quad \text{if } \alpha > 1. \end{aligned} \quad (4.43)$$

The Mean of  $W^{(2)}$

In a group of  $2^L$  sources, we set  $W_R|k$  to be the sum of the passive waiting steps of the accepted packets during an epoch of  $k$  steps. We have

$$\begin{aligned} E[W_R|k] &= 2^L \sum_{i=1}^k p q^{i-1}(k-i) \\ &= 2^L(k - \frac{1-q^k}{p}). \end{aligned} \quad (4.44)$$

Set  $W_n^{(2)}$  to be the sum of the passive waiting steps of the accepted packets during the  $n^{\text{th}}$  epoch and  $N_n$  to be the total number of packets transmitted during the  $n^{\text{th}}$  epoch. We have

$$E[W^{(2)}] = \frac{\sum_{n=1}^{\infty} W_n^{(2)}}{\sum_{n=1}^{\infty} N_n}, \text{ W.P.1.} \quad (4.45)$$

By applying a Markov Ratio Limit Theorem (see [ 10 ], [ 24 ] and [ 28 ]), we have

$$E[W^{(2)}] = \frac{\sum_{j=0}^{2^L} \pi_j \sum_{k=1}^{\infty} P_{T_L|j}(k|j) E[W_R|k]}{\sum_{j=0}^{2^L} j \pi_j}, \text{ W.P.1.} \quad (4.46)$$

Incorporate the result of  $E[W_R|k]$  given by Equation (4.44) into Equation (4.46), we have

$$E[W^{(2)}] = \frac{\sum_{j=0}^{2^L} \pi_j \sum_{k=1}^{\infty} P_{T_L|j}(k|j) (k - \frac{1-q^k}{p})}{\sum_{j=0}^{2^L} j \pi_j}, \text{ W.P.1.} \quad (4.47)$$

### The Mean of $W^{(3)}$

Set  $N_n$  to be the number of packets transmitted during the  $n^{\text{th}}$  epoch and  $W_n^{(3)}$  to be the sum of the active waiting steps of the packets transmitted during the  $n^{\text{th}}$  epoch. We have

$$E[W^{(3)}] = \frac{\sum_{n=1}^{\infty} W_n^{(3)}}{\sum_{n=1}^{\infty} N_n}, \text{ W.P.1.} \quad (4.48)$$

By applying the Markov Ratio Limit Theorem, we have

$$E[W^{(3)}] = \frac{\sum_{j=0}^{2^L} E[W_L|j] \pi_j}{\sum_{j=0}^{2^L} j \pi_j}, \text{ W.P.1.} \quad (4.49)$$

We are now ready to calculate the performance measures of the systems implemented by Policy 1 and Policy 2.

### Policy 1

The system's average packet input rate is given by Equation (4.36).

Set  $S_n$  to be the number of slots in the  $n^{\text{th}}$  epoch and  $N_n$  to be the number of packets transmitted during the  $n^{\text{th}}$  epoch. By definition, we have

$$S = \frac{\sum_{n=1}^{\infty} N_n}{\sum_{n=1}^{\infty} S_n}, \text{ W.P.1.} \quad (4.50)$$

By applying the Markov Ratio Limit Theorem, we have

$$S = \frac{\sum_{j=0}^{2^L} j \pi_j}{\sum_{j=0}^{2^L} \sum_{k=0}^{\infty} k P_{T_L|j}(k|j) \pi_j}, \text{ W.P.1.} \quad (4.51)$$

Obviously, the loss probability of a new arriving packet is

$$\ell = 1 - \frac{S}{N_T E[A]}. \quad (4.52)$$

From Equation (4.1), the average delay of the accepted packets is

$$E[D] = E[W_S^{(1)}] + (E[W^{(2)}] + E[W^{(3)}])\beta + 1 + R. \quad (4.53)$$

Where  $E[W_S^{(1)}]$ ,  $E[W^{(2)}]$  and  $E[W^{(3)}]$  are given by Equations (4.43), (4.47) and (4.49), respectively.

## Policy 2

The system's average packet input rate is given by Equation (4.36).

By similar techniques as under Policy 1, we have

$$S = \frac{\sum_{j=0}^{2^L} j \pi_j}{2 \sum_{j=0}^{2^L} \sum_{k=0}^{\infty} k P_{T_L|j}(k|j) \pi_j}, \text{ W.P.1.} \quad (4.54)$$

The loss probability of a new arriving packet is

$$\ell = 1 - \frac{S}{N_T E[A]}. \quad (4.55)$$

The average delay of the accepted packets is

$$E[D] = E[W_S^{(1)}] + (E[W^{(2)}] + E[W^{(3)}])\beta + 1.5 + R, \quad (4.56)$$

where  $E[W_S^{(1)}]$ ,  $E[W^{(2)}]$  and  $E[W^{(3)}]$  are given by Equations (4.43), (4.47), and (4.49), respectively. Note that because of the additional delay of one slot for every packet successfully transmitted in the second slot of a step, by symmetry, the average successful transmission delay of the accepted packets is 1.5 slot.

## A Fixed-Assignment TDMA Scheme

All along, we have assumed that the messages arriving at the sources are single packets. Let's now relax this constraint when the system is implemented by a Fixed-Assignment TDMA scheme. The messages arriving at the sources are now assumed to be composed of multiple packets. Each buffer space is assumed to be able to hold one message. The numbers of packets composing these messages are assumed to be



independent and identically distributed. Let  $B$  be the number of packets in a message. Set  $P_B(i)$  to be the distribution of  $B$  and  $C_B(z)$  to be the  $z$ -transform of  $P_B(i)$ . By definitions, we have

$$P_{T_0|j}(k|j) = \begin{cases} \delta_{1k} & , \text{ for } j = 0 , \\ P_B(k) & , \text{ for } j = 1 , \end{cases} \quad (4.57)$$

and

$$E[W_0|j] = \begin{cases} 0 & , \text{ for } j = 0 , \\ E[B]-1 & , \text{ for } j = 1 , \end{cases} \quad (4.58)$$

where

$$\delta_{1k} = \begin{cases} 1 , & \text{ if } k = 1 , \\ 0 , & \text{ otherwise .} \end{cases}$$

By applying Theorem 4.2, we have

$$P_j = \begin{cases} P & , \text{ for } j = 0 , \\ 1 - C_B(q) & , \text{ for } j = 1 . \end{cases} \quad (4.59)$$

and

$$\Pi_0 = \frac{C_B(q)}{P + C_B(q)} , \quad (4.60)$$

$$\Pi_1 = \frac{P}{P + C_B(q)} . \quad (4.61)$$

The system's average message input rate is

$$\lambda = \alpha E[A] . \quad (4.62)$$

By applying the Markov Ratio Limit Theorem, the loss probability of a new arriving message is given by

$$\rho = 1 - \frac{\pi_1}{\alpha E[A] (\pi_0 + \pi_1 E[B])} . \quad (4.63)$$

The average frame latency of the accepted messages is given by Equation (4.43).

By applying Equation (4.47) , we have

$$E[W^{(2)}] = E[B] - \frac{1}{P} (1 - C_B(q)) . \quad (4.64)$$

By applying Equation (4.49), we have

$$E[W^{(3)}] = E[B] - 1 . \quad (4.65)$$

By applying Equation (4.51), we have

$$S = \frac{\pi_1 E[B]}{\pi_0 + \pi_1 E[B]} . \quad (4.66)$$

Under a Fixed-Assignment TDMA scheme, no collisions occur and the construction of superframes is not necessary. Hence,  $\alpha = \beta$ , and the average delay of an accepted message is given by

$$E[D] = W_S^{(1)} + (E[W^{(2)}] + E[W^{(3)}])\alpha + 1 + R , \quad (4.67)$$

where  $E[W_S^{(1)}]$ ,  $E[W^{(2)}]$  and  $E[W^{(3)}]$  are given by Equations (4.43), (4.64) and (4.65), respectively.

### Numerical Results

To demonstrate the performance of the system implemented by Policy 1, we consider a population of 64 sources. The arrival process of the packets at a source is described by the Bernuolli trial distribution

$$P_A(i) = \begin{cases} 1-P_S, & \text{if } i = 0. \\ P_S, & \text{if } i = 1, \\ 0, & \text{otherwise.} \end{cases}$$

A family of access-control schemes is obtained by dividing, according to different values of  $L$ , the entire population into groups of an equal number of sources as described in Section 4.2. The possible values of  $L$  are 0, 1, 2, 3, 4, 5, 6. Hence, there are seven access-control schemes in this family, each associated with an appropriate value of  $L$ . The joint performance of this family of access-control schemes is demonstrated by curves plotted in Figures 4.8-4.12. Note that when  $L = 0$ , it is a Fixed-Assignment TDMA scheme. When  $L = 6$ , it is a pure Tree Search scheme.

Figure 4.8 plots the mean of  $W^{(2)} + W^{(3)}$  of the packets versus the throughput,  $S$ , of the system. From the figure, it is evident that a scheme of higher  $L$  value on the average encounters more collisions per epoch, and thus has longer average epoch length than does a scheme of lower  $L$  value. When the average packet input rate of the system is increased, the throughput,  $S$ , of the system is also increased. However, when the average packet input rate is increased beyond a certain limit, frequent collisions of packets not only increase the mean of  $W^{(2)} + W^{(3)}$  but also reduce the throughput,  $S$ , of the system,

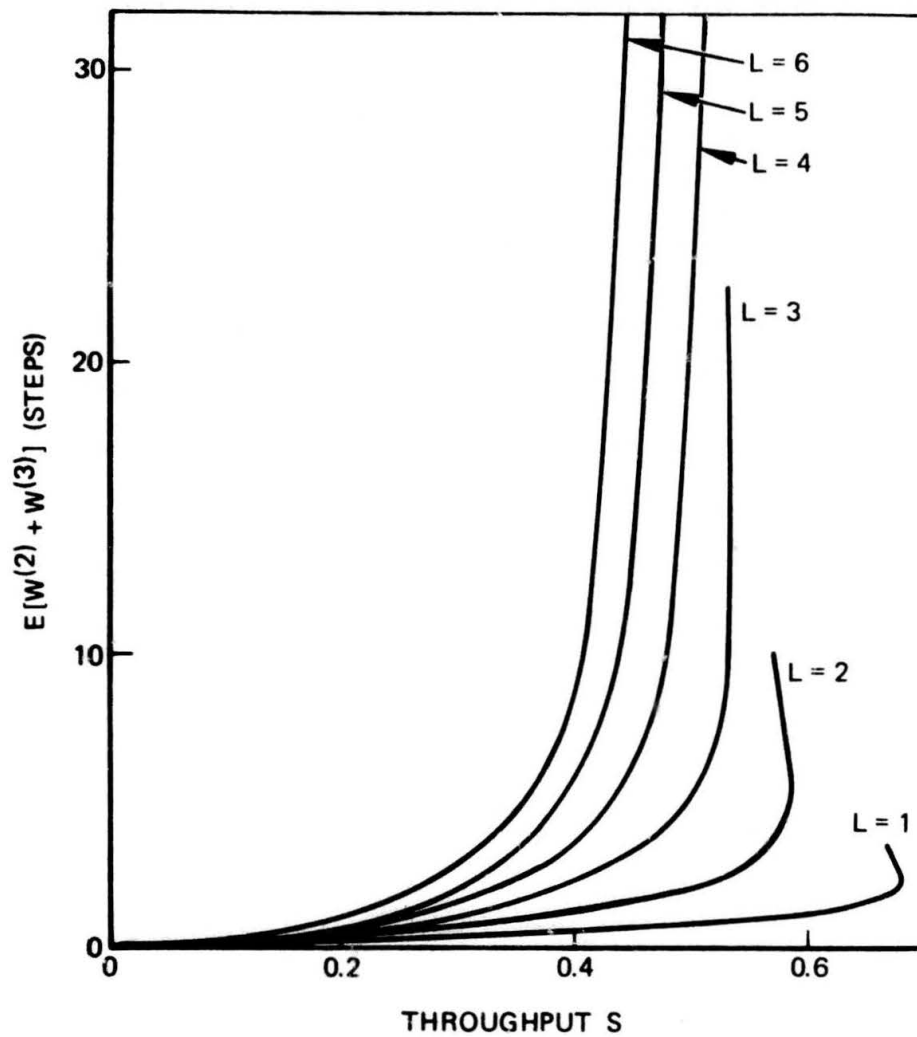


Figure 4.8.  $E[W^{(2)} + W^{(3)}]$  vs Throughput Curves under Policy 1.

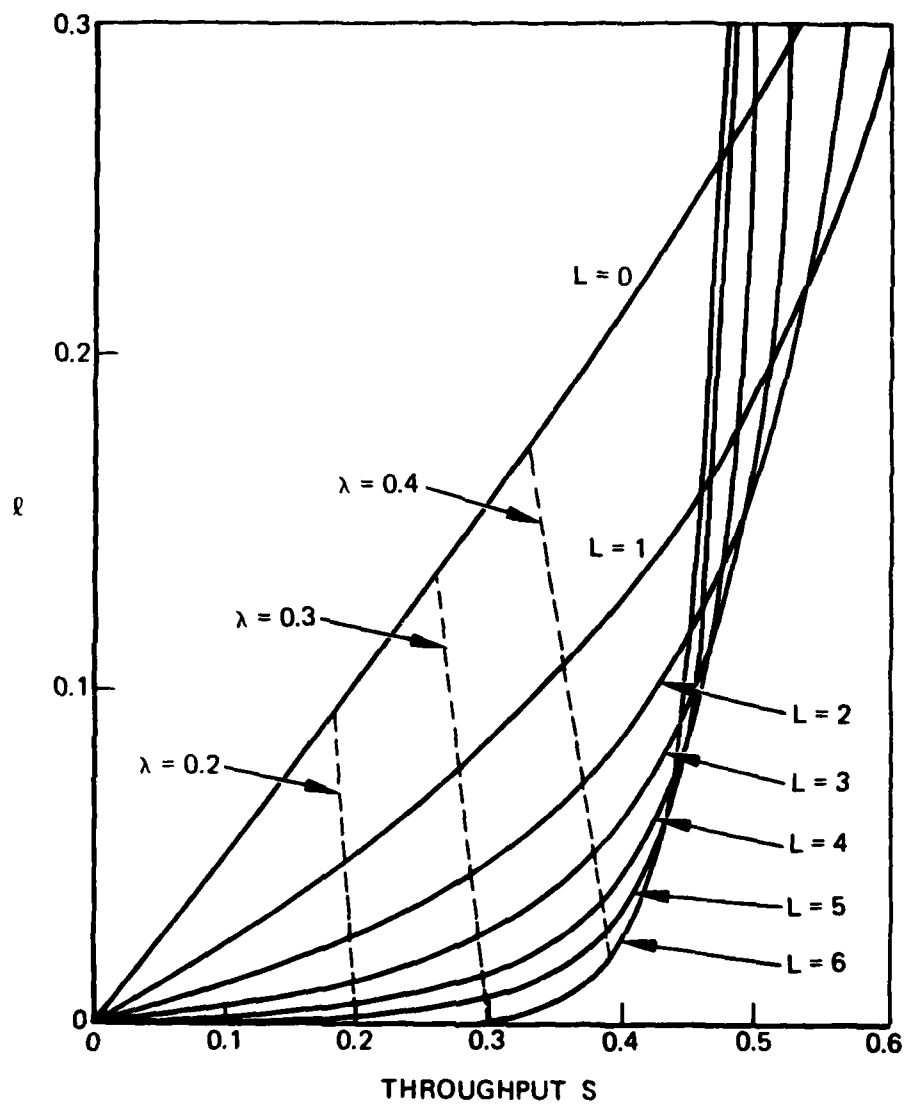


Figure 4.9.  $\rho$  vs Throughput Curves under Policy 1.

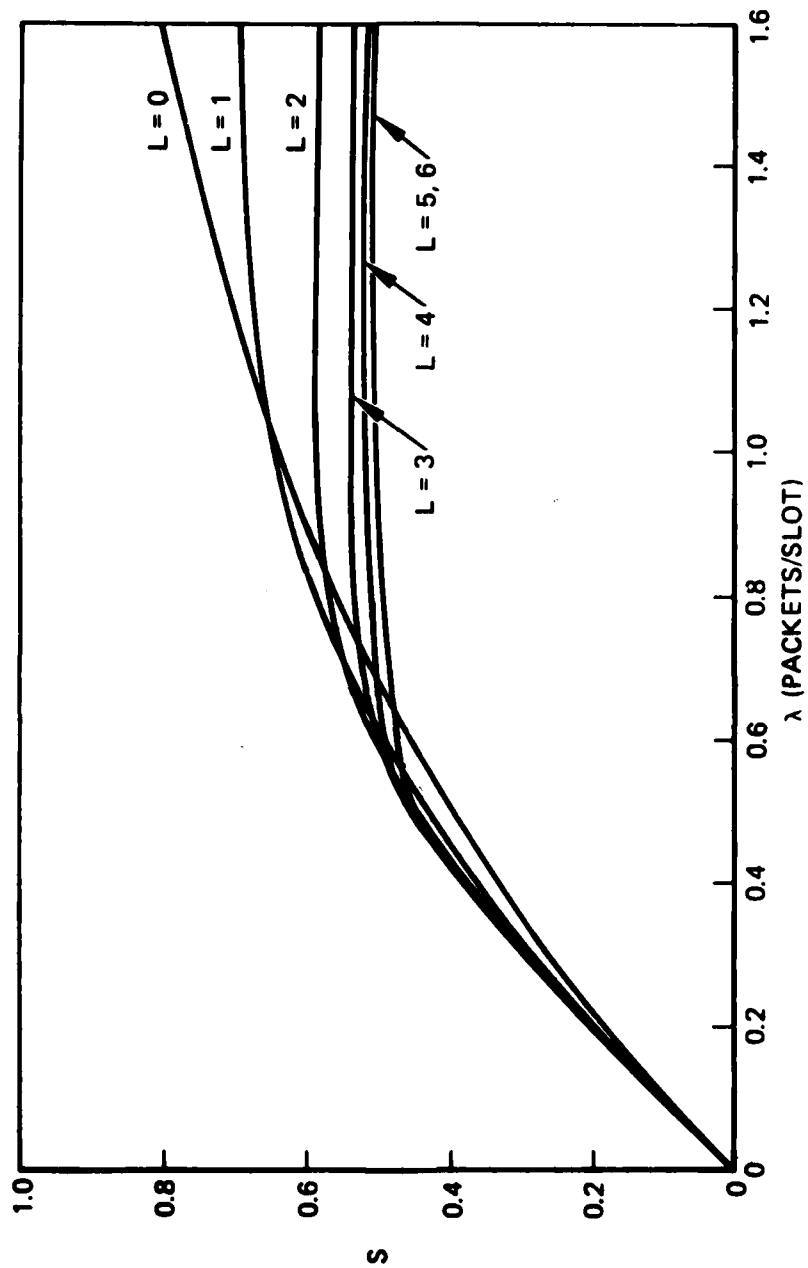


Figure 4.10. Throughput vs  $\lambda$  Curves under Policy 1.

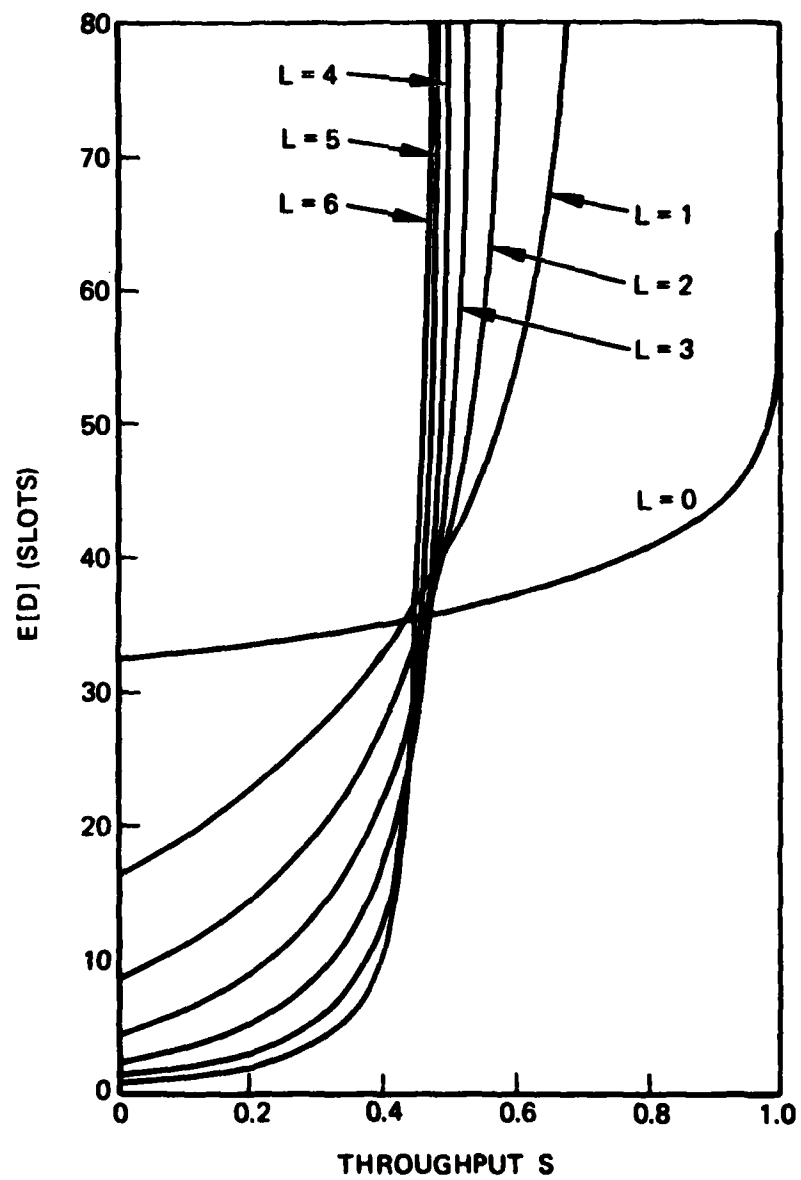


Figure 4.11. Delay – Throughput Curves under Policy 1,  $R = 0$ .

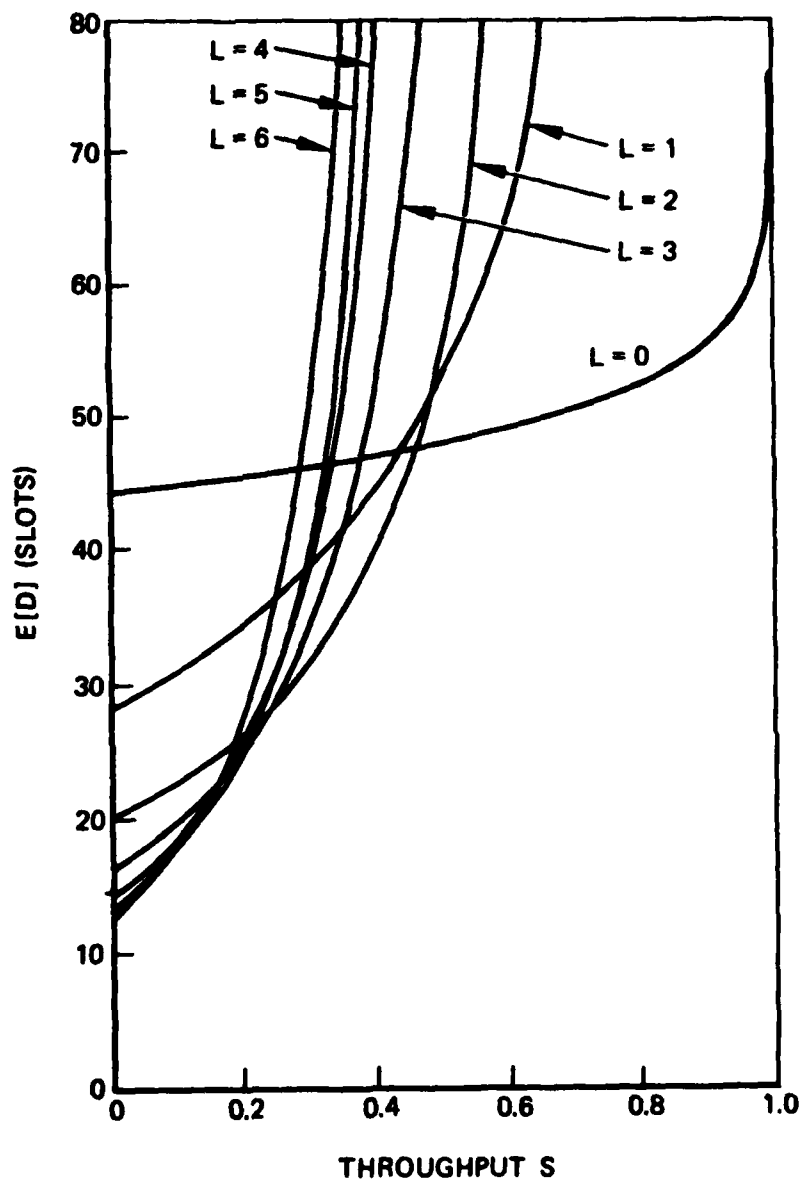


Figure 4.12. Delay - Throughput Curves under Policy 1,  $R = 12$ .



This is well demonstrated by the curve  $L = 1$ .

Figure 4.9 plots the loss probability,  $\ell$ , of a new arriving packet versus the throughput,  $S$ , of the system. A scheme of higher  $L$  value has more sources per group and shorter frame duration than does a scheme of lower  $L$  value. Hence, if the throughput,  $S$ , is low, we obtain a lower value of  $\ell$  when the system is implemented by a scheme of higher  $L$  value than when the system is implemented by a scheme of lower  $L$  value. But if the throughput,  $S$ , is increased beyond a certain value, the effect of frequent packet collisions prevails and the opposite is true. The same reason applies in Figure 4.10, which plots the throughput,  $S$ , of the system versus its average packet input rate  $\lambda$ .

Figures 4.11 and 4.12 plot the average delay,  $E[D]$ , of the accepted packets versus the throughput,  $S$ , of the system. When the throughput,  $S$ , is low, the frame latency,  $W_S^{(1)}$  is the dominant part of the delay of an accepted packet. A scheme of higher  $L$  value has shorter-frame duration and yields lower average delay of the accepted packets than does a scheme of lower  $L$  value. But if the throughput,  $S$ , of the system is increased beyond a certain value, because of frequent packet collisions,  $W_S^{(2)} + W_S^{(3)}$  prevails over  $W_S^{(1)}$  and we have opposite results. Notice that because of the construction of superframes, if the throughput,  $S$ , is low, the curves of high  $L$  values in Figure 4.12 ( $R = 12$ ) are much steeper than the curves of high  $L$  values in Figure 4.11 ( $R = 0$ ). Also notice that in Figure 4.11 ( $R = 0$ ), if the throughput,  $S$ , is lower than 0.45, the optimal value of  $L$  is 0 and if the throughput,  $S$ , is higher than 0.45, the optimal value of  $L$  is 6. Values of  $L$  beside 0 and 6 are optimal only when the throughput  $S$ , is

in a very small vicinity of 0.45. This is not so in Figure 4.12 ( $R = 12$ ) when the construction of superframes is necessary.

When the system is implemented by Policy 2, possible values of  $L$  are 1, 2, 3, 4, 5, 6. Hence, there are six access-control schemes in this family. The joint performance of this family of access-control schemes is demonstrated by curves plotted in Figures 4.13-4.17. They look similar to the performance curves of the system when implemented by Policy 1. Note that when  $L = 1$ , it is a Fixed-Assignment TDMA scheme. When  $L = 6$ , it is a pure Tree Search scheme.

It is observed that if the throughput,  $S$ , of the system is low, the system performs better both in terms of the loss probability of a new arriving packet and the average delay of the accepted packets when implemented by Policy 1 than when implemented by Policy 2. If the throughput,  $S$ , of the system is high, the opposite is true.

To demonstrate the effect of propagation delay on the average delay of the accepted packets, we plot, in Figure 4.18, the lower envelopes of the curves in Figures 4.12 and 4.17. The resulting curves show better performance of the system when implemented by Policy 2 than when implemented by Policy 1, unless the throughput,  $S$ , is very low. This is consistent with what we suggested in Section 4.3.

#### 4.6 Sources with Unlimited Buffer Capacities

In this section, we will study the case in which the source buffers are assumed to have unlimited capacities. First, we describe the source buffers and the statistics of the arrivals of packets at each source. Then, we analyze the systems implemented by Policy 1 and Policy 2.

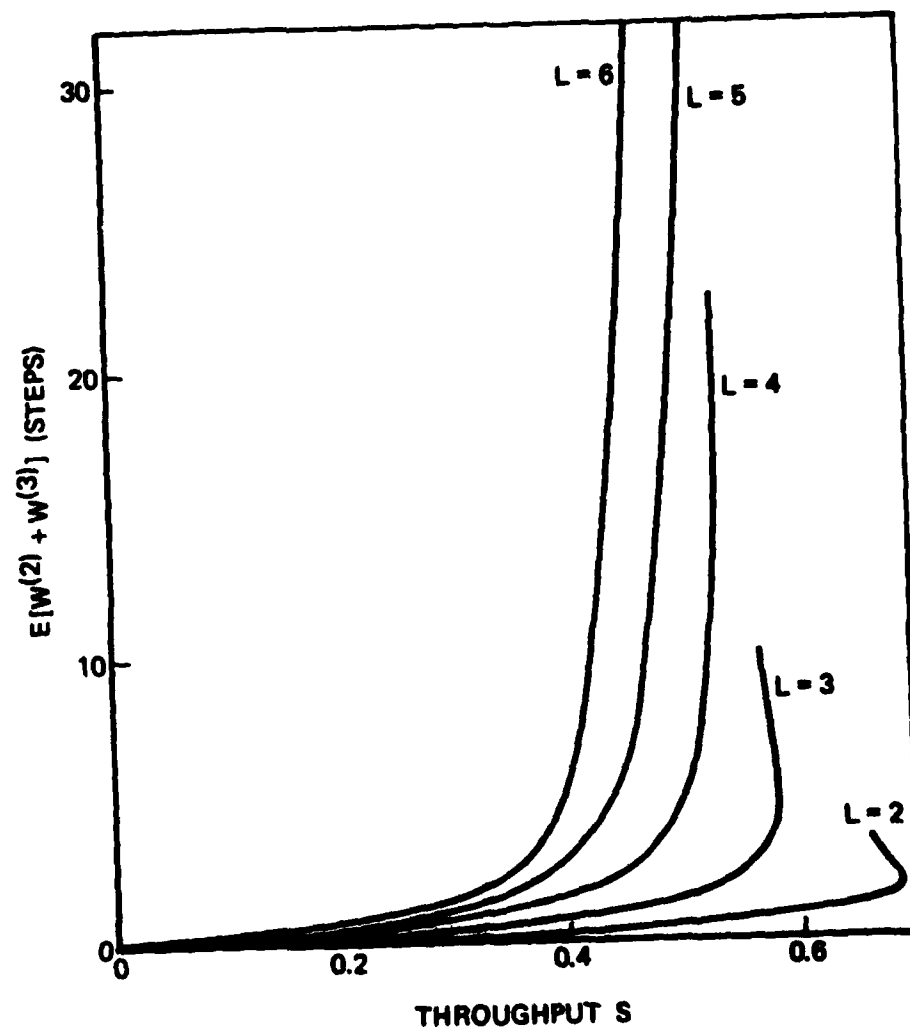


Figure 4.13.  $E[W^{(2)} + W^{(3)}]$  vs Throughput Curves under Policy 2.

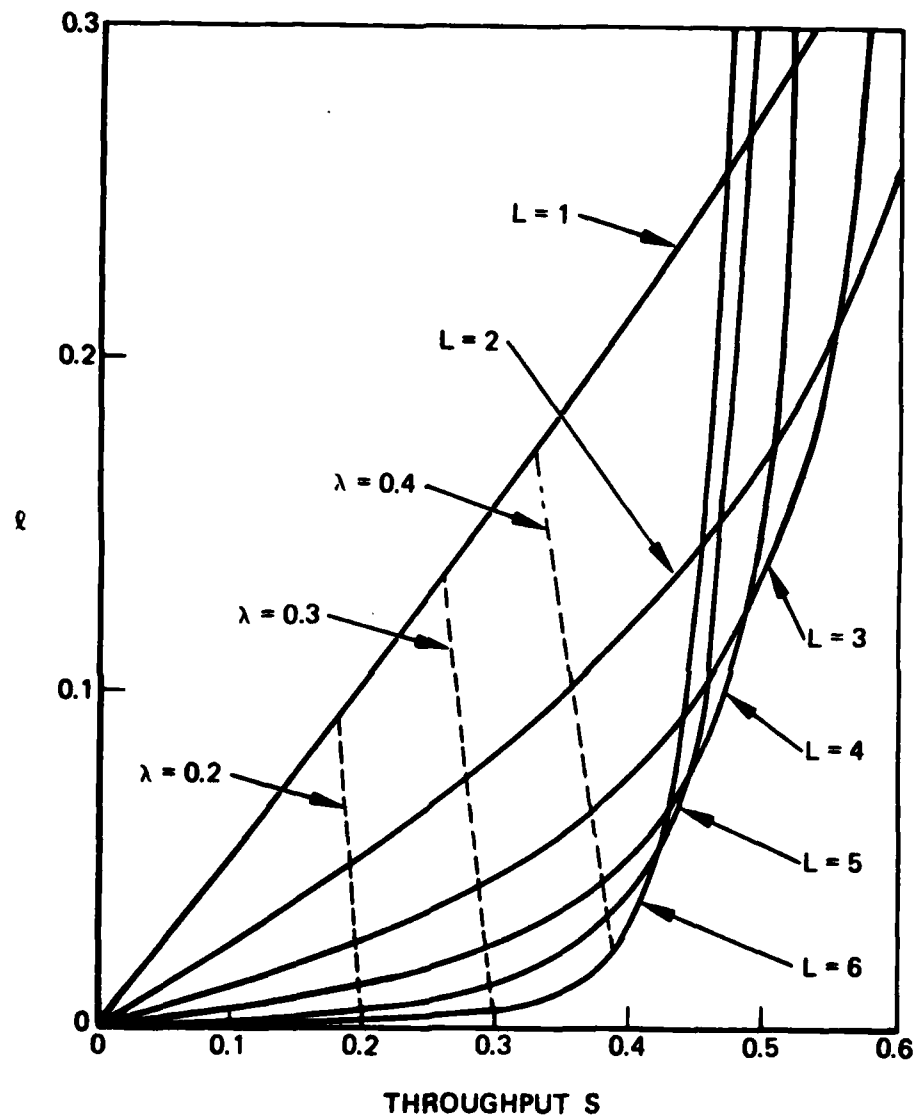


Figure 4.14.  $l$  vs Throughput Curves under Policy 2.

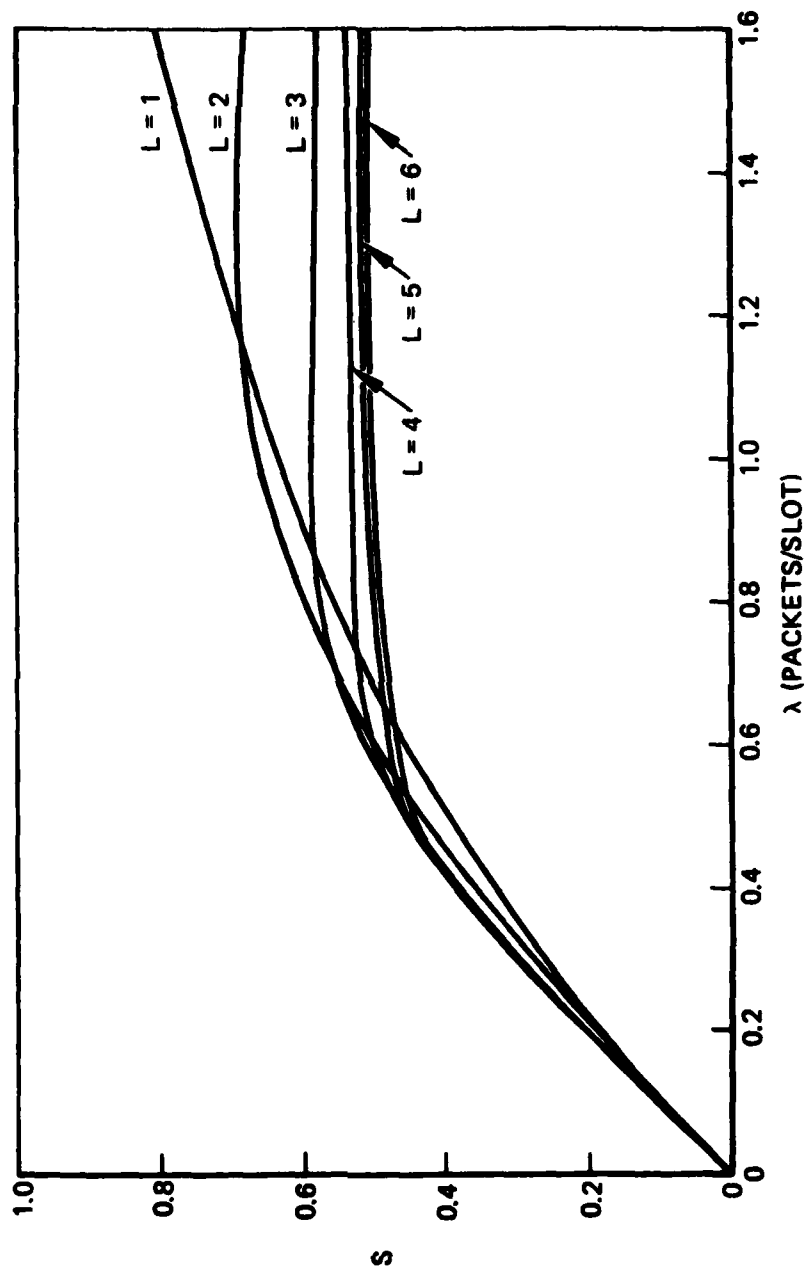


Figure 4.15. Throughput vs  $\lambda$  Curves under Policy 2.

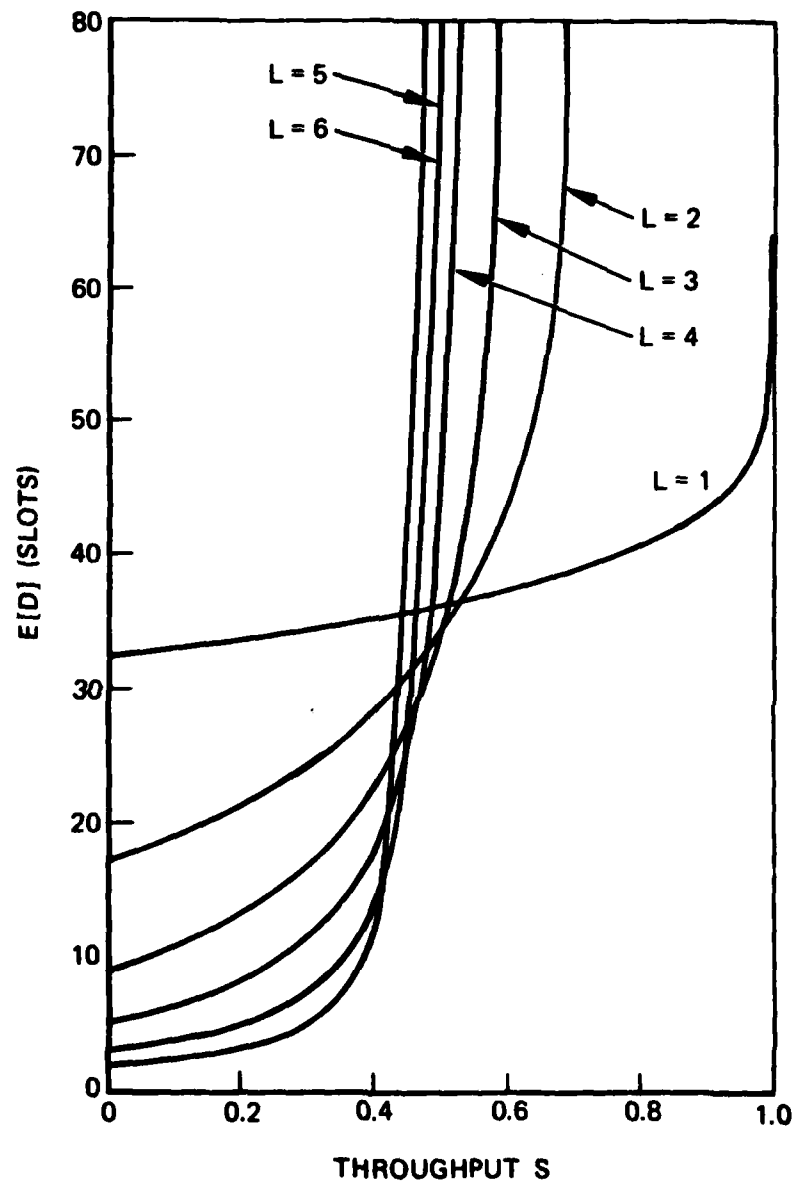


Figure 4.16. Delay - Throughput Curves under Policy 2,  $R = 0$ .

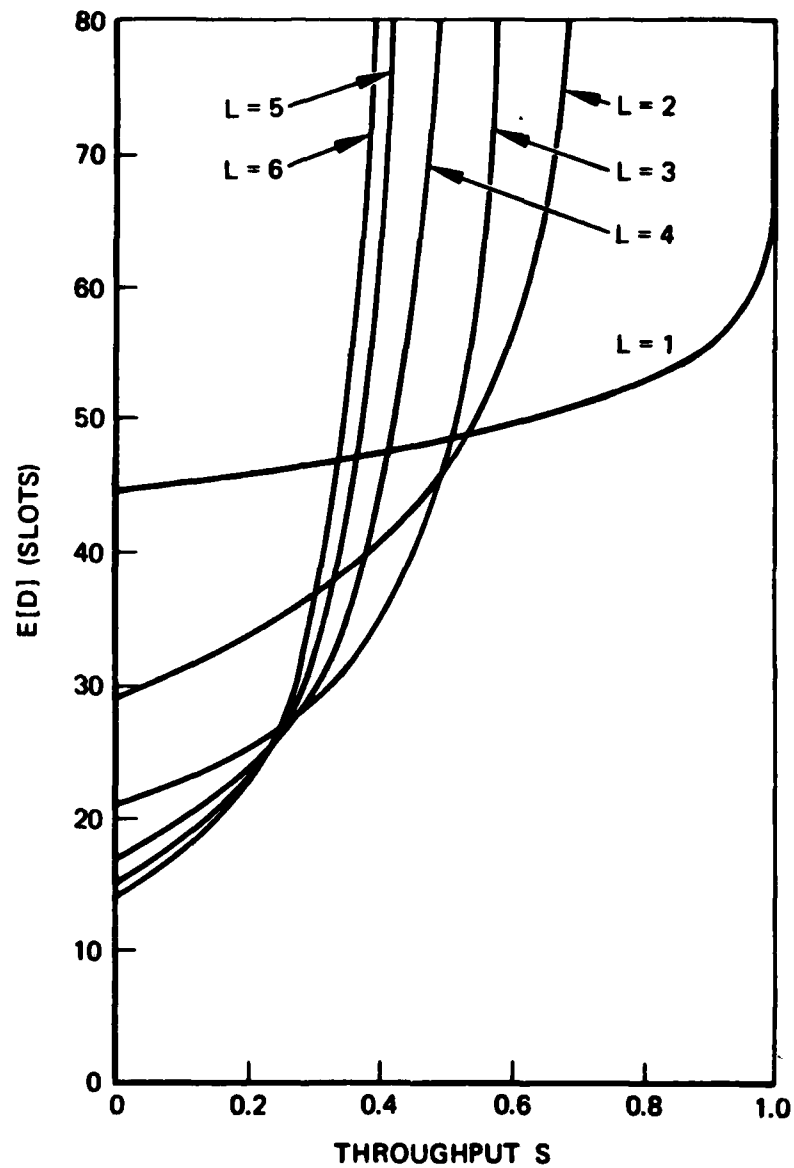


Figure 4.17. Delay - Throughput Curves under Policy 2,  $R = 12$ .

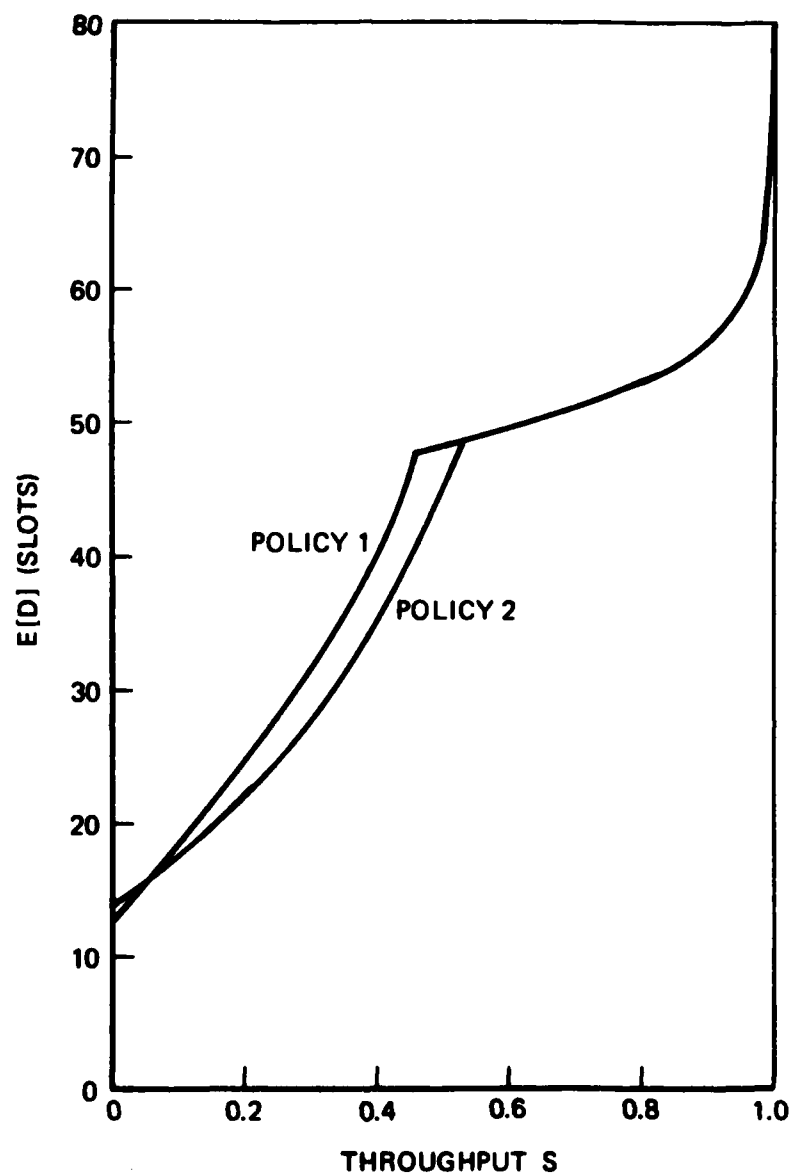


Figure 4.18. Delay - Throughput Curves (Lower Envelopes) under Policy 1 and Policy 2,  $R = 12$ .



### The Source Model

Again, by symmetry, we consider only one group of sources and the slots assigned to this group. The arrival process of packets at each source is again described by  $A$ ,  $P_A(1)$ , and  $C_A(z)$  defined at the beginning of Section 4.5. However, the buffer capacity of each source is now assumed to be unlimited. Packets arriving at each source are transmitted in a first come first served order.

Set  $\underline{Y}_n$ ,  $n > 0$ , to be the joint queue size of all the queues at the sources at the beginning of the  $n^{\text{th}}$  epoch, then  $\underline{Y}_n$ ,  $n > 0$ ,  $\underline{Y}_n = Y_n^{(1)}, Y_n^{(2)}, \dots, Y_n^{(G_S)}$ , is over the state space  $\mathcal{S} = \prod_{j=1}^{G_S} I_j$ , where  $I_j$  is the set of non-negative integers. The development of the  $n^{\text{th}}$  epoch depends only on  $\underline{Y}_n$ ,  $n > 0$ . Also, the numbers of packets arriving at all the sources during the  $n^{\text{th}}$  epoch depend only on the duration of the epoch. Hence, the future development of the process depends only on  $\underline{Y}_n$ ,  $n > 0$ . Consequently, the sequence  $\{\underline{Y}_n, n > 0\}$  describes the state of the process. Therefore,  $\{\underline{Y}_n, n > 0\}$  is a Markov Chain.

A complete description of this Markov Chain means the specification of joint probabilities depending upon  $2^L$  indices. Each indice corresponds to the queue of one source. Hence, the determination of these probabilities demands the solution of a large number of sets of equations of complicated form that no simple useful analytical expressions for this Markov Chain can be obtained. Consequently, we develop the following approximation of the joint queue size distribution of this Markov Chain.

### Approximation of the Joint Queue Size Distribution

In this joint queue size distribution approximation (see [29 ], [30 ], and [31 ]), we assume the same distribution of packets at each source meets the group slot(s) at the beginning of each epoch. Then we can find the distribution of packets at each source at the beginning of an epoch by the following argument. If the group slot(s) meets the same distribution of packets at each source at the beginning of an epoch, then the group slot(s) must meet the same distribution of packets at each source at the beginning of the next epoch. By symmetry, we need to study the queue size distribution of packets at only one specific source. Henceforth, we call this source Source A. We make the following definitions with respect to Source A:

$F_i$  - the probability of having  $i$  packets,  $i \geq 0$ , at the beginning of an epoch at steady state, if it exists.

$F_{>0}$  - the probability of having at least one packet at the beginning of an epoch at steady state, if it exists.  $F_{>0} = 1 - F_0$ .

$F_{in}$  - the probability of having  $n$  packets at the beginning of the next epoch given there are  $i$  packets at the beginning of the present epoch.

$U_0$  - the duration of an epoch given Source A has no packet at the beginning of the epoch.

$P_{U_0}(j)$  - the distribution of  $U_0$ .

$C_{U_0}(z)$  - the z-transform of  $P_{U_0}(j)$ .

$U_{>0}$  - the duration of an epoch given Source A has at least one packet at the beginning of the epoch.

$P_{U_{>0}}(j)$  - the distribution of  $U_{>0}$ .

$C_{U_{>0}}(z)$  - the z-transform of  $P_{U_{>0}}(j)$ .

### Analysis of the Systems

Let  $i$  be the number of sources, beside Source A, which have packets to transmit during an epoch. Since the distributions of packets at all the sources at the beginning of an epoch are statistically independent, the random variable  $i$  assumes the Binomial distribution

$$\binom{2^L-1}{i} F_{>0}^i F_0^{2^L-1-i}.$$

Recall that a source can transmit at most one packet during an epoch.

Hence, we have

$$P_{U_0}(k) = \sum_{i=0}^{2^L-1} \binom{2^L-1}{i} F_{>0}^i F_0^{2^L-1-i} P_{T_L|i}(k|i), \quad (4.68)$$

and

$$C_{U_0}(z) = \sum_{i=0}^{2^L-1} \binom{2^L-1}{i} F_{>0}^i F_0^{2^L-1-i} C_{T_L|i}(z|i), \quad (4.69)$$

where  $C_{T_L|i}(z|i)$  is the z-transform of  $P_{T_L|i}(k|i)$ .

Similarly, we have

$$P_{U_{>0}}(k) = \sum_{i=0}^{2^L-1} \binom{2^L-1}{i} F_{>0}^i F_0^{2^L-1-i} P_{T_L|i+1}(k|i+1), \quad (4.70)$$

and

$$C_{U>0}(z) = \sum_{i=0}^{2^L-1} \binom{2^L-1}{i} F_{>0}^i F_0^{2^L-1-i} P_{T_L|i+1}(z|i+1), \quad (4.71)$$

where  $C_{T_L|i+1}(z|i+1)$  is the z-transform of  $P_{T_L|i+1}(k|i+1)$ .

#### Theorem 4.3

Assume a system implemented by a Policy 1 or Policy 2 scheme with  $2^L$  sources in each group. Set  $Y_n$ ,  $n > 0$ , to be the number of packets at a source at the beginning of the  $n^{\text{th}}$  epoch. Under the joint queue size distribution approximation,  $\{Y_n, n > 0\}$  is a Markov Chain over the state space of non-negative integers. If  $P_A(0) < 1$ , this Markov Chain is irreducible and aperiodic with state transition distribution given by

$$F_{0n} = \begin{cases} \sum_{k=0}^{\infty} P_A^{(k\alpha)}(n) P_{U_0}(k), & 0 \leq n < \infty, \\ 0 & \text{otherwise,} \end{cases} \quad (4.72)$$

and

$$F_{in} = \begin{cases} \sum_{k=0}^{\infty} P_A^{(k\alpha)}(n+1-i) P_{U_{>0}}(k), & i-1 \leq n < \infty \\ 0 & \text{otherwise,} \end{cases} \quad (4.73)$$

where  $P_A^{(k\alpha)}(n)$  is the  $k\alpha$ -time convolution of  $P_A(n)$ .

This Markov Chain is positive recurrent iff

$$0 < E[A] < \frac{1}{\alpha E[U_{>0}]} \quad (4.74)$$

Furthermore, if this Markov Chain is positive recurrent, the z-transform of its steady state distribution is given by

$$F(z) = \frac{F_0(C_{U_0}(\phi) - \frac{C_{U_{>0}}(\phi)}{z})}{1 - \frac{C_{U_{>0}}(\phi)}{z}}, \quad (4.75)$$

where

$$\phi = C_A(z)^\alpha, \quad (4.76)$$

and

$$F_0 = \frac{1 - \alpha E[A]E[U_{>0}]}{1 + \alpha E[A]E[U_0] - \alpha E[A]E[U_{>0}]}. \quad (4.77)$$

#### Proof

By definition,  $\{Y_n, n > 0\}$  is a Markov Chain.

Also by definition, Equations (4.72) and (4.73) are obtained.

If  $P_A(0) < 1$ , every state can be reached from every other state with non-zero probability given by Equations (4.72) and (4.73). Hence, the Markov Chain  $\{Y_n, n > 0\}$  is irreducible and aperiodic.

Assume the Markov Chain  $\{Y_n, n > 0\}$  is positive recurrent, then the steady state distribution of  $\{Y_n, n > 0\}$  is the unique solution of the following set of equations:

$$F_n = \sum_{i=0}^{\infty} F_i F_{in}, \quad 0 \leq n < \infty, \quad (4.78)$$

and

$$\sum_{n=0}^{\infty} F_n = 1, \quad F_n \geq 0, \quad 0 \leq n < \infty. \quad (4.79)$$

Taking z-transform of Equation (4.78), we have

$$\begin{aligned}
 F(z) &= \sum_{n=0}^{\infty} F_n z^n \\
 &= \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} F_i F_{in} z^n \\
 &= F_0 \sum_{n=0}^{\infty} F_{0n} z^n + \sum_{i=1}^{\infty} F_i \sum_{n=0}^{\infty} F_{in} z^n .
 \end{aligned} \tag{4.80}$$

Consider the first term at the right hand side of Equation (4.80).

From Equation (4.72), we have

$$\begin{aligned}
 F_0 \sum_{n=0}^{\infty} F_{0n} z^n &= F_0 \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} P_A^{(k\alpha)}(n) P_{U_0}(k) z^n \\
 &= F_0 \sum_{k=0}^{\infty} P_{U_0}(k) \sum_{n=0}^{\infty} P_A^{(k\alpha)}(n) z^n \\
 &= F_0 \sum_{k=0}^{\infty} P_{U_0}(k) C_A(z)^{k\alpha} \\
 &= F_0 C_{U_0}(C_A(z)^\alpha) .
 \end{aligned} \tag{4.81}$$

Consider the second term at the right hand side of Equation (4.80).

From Equation (4.73), we have

$$\begin{aligned}
 \sum_{i=1}^{\infty} F_i \sum_{n=0}^{\infty} F_{in} z^n &= \sum_{i=1}^{\infty} F_i \sum_{n=i-1}^{\infty} \sum_{k=0}^{\infty} P_A^{(k\alpha)}(n+1-i) P_{U_{>0}}(k) z^n \\
 &= \sum_{i=1}^{\infty} F_i \sum_{k=0}^{\infty} P_{U_{>0}}(k) \sum_{n=i-1}^{\infty} P_A^{(k\alpha)}(n+1-i) z^n \\
 &= \sum_{i=1}^{\infty} F_i \sum_{k=0}^{\infty} P_{U_{>0}}(k) z^{i-1} \sum_{n=i-1}^{\infty} P_A^{(k\alpha)}(n+1-i) z^{n+1-i}
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{\infty} F_i z^{i-1} \sum_{k=0}^{\infty} P_{U>0}(k) C_A(z)^{k\alpha} \\
&= \sum_{i=1}^{\infty} F_i z^{i-1} C_{U>0}(C_A(z)^\alpha) .
\end{aligned} \tag{4.82}$$

Set  $\phi = C_A(z)^\alpha$ .

From Equation (4.81), we have

$$F_0 \sum_{n=0}^{\infty} F_{0n} z^n = F_0 C_{U_0}(\phi). \tag{4.83}$$

From Equation (4.82), we have

$$\sum_{i=1}^{\infty} F_i \sum_{n=0}^{\infty} F_{in} z^n = \sum_{i=1}^{\infty} F_i z^{i-1} C_{U>0}(\phi). \tag{4.84}$$

Incorporating Equations (4.83) and (4.84) into Equation (4.80), we have

$$\begin{aligned}
F(z) &= F_0 C_{U_0}(\phi) + \sum_{i=1}^{\infty} F_i z^{i-1} C_{U>0}(\phi) \\
&= F_0 C_{U_0}(\phi) + C_{U>0}(\phi) \frac{1}{z} \sum_{i=1}^{\infty} F_i z^i \\
&= F_0 C_{U_0}(\phi) + C_{U>0}(\phi) \frac{1}{z} (F(z) - F_0) \\
&= \frac{F_0(C_{U_0}(\phi) - \frac{C_{U>0}(\phi)}{z})}{1 - \frac{C_{U>0}(\phi)}{z}} .
\end{aligned}$$

Evaluate  $F(z)$  at  $z = 1$ , we have

$$F(z) \Big|_{z=1} = \frac{F_0 \{E[U_0] \alpha E[A] - (E[U_{>0}] \alpha E[A] - 1)\}}{-(E[U_{>0}] \alpha E[A] - 1)}$$

Since  $F(1) = 1$ , we have

$$\frac{F_0 \{E[U_0] \alpha E[A] - (E[U_{>0}] \alpha E[A] - 1)\}}{-(E[U_{>0}] \alpha E[A] - 1)} = 1$$

Rearrange terms, Equation (4.77) is obtained.

The Markov Chain  $\{Y_n, n > 0\}$  is positive recurrent iff  $1 > F_0 > 0$  (see [28] and [29]). From Equation (4.77), this condition is satisfied iff

$$1 - \alpha E[A] E[U_{>0}] > 0,$$

which is equivalent to Condition (4.74).

Q.E.D.

#### Corollary 4.1

Consider the Markov Chain  $\{Y_n, n > 0\}$  defined in Theorem 4.3. Set  $\tilde{Y}_n = Y_n - 1 | Y_n > 0, n > 0$ , which is the number of packets at the source at the beginning of the  $n^{\text{th}}$  epoch excluding the packet transmitted during the epoch given that there is at least one packet at the beginning of the  $n^{\text{th}}$  epoch. Under the joint queue size distribution approximation, if Condition (4.74) is satisfied, the steady state distribution of this embedded queue, denoted by  $\{H_i, i \geq 0\}$ , is given by

$$H_i = \frac{F_{i+1}}{F_{>0}}. \quad (4.85)$$



The z-transform of  $\{H_i, i \geq 0\}$ , denoted by  $H(z)$ , is given by

$$H(z) = \frac{F_0}{F_{>0}} \frac{C_{U_0}(\phi) - 1}{z - C_{U_{>0}}(\phi)} \quad (4.86)$$

Furthermore, the mean of this imbedded queue, denoted by  $E[Q]$ , is given by

$$E[Q] = \frac{\alpha E[A]^2 E[U_0^2] + E[U_0](\text{Var}(A) - E[A])}{2E[A]E[U_0]} + \frac{\alpha^2 E[A]^2 E[U_{>0}^2] + \alpha E[U_{>0}](\text{Var}(A) - E[A])}{2(1 - \alpha E[A]E[U_{>0}])} \quad (4.87)$$

where  $\text{Var}(A)$  is the variance of  $A$ .

#### Proof

The distribution  $\{H_i, i \geq 0\}$  is obtained from the distribution  $\{F_i, i \geq 0\}$ .  $H(z)$  is deduced from  $F(z)$  given by Theorem 4.3.  $E[Q]$  is obtained by differentiating  $H(z)$  with respect to  $z$  and then evaluate it at  $z = 1$ .

Q.E.D.

#### Theorem 4.4

Assume a system implemented by a Policy 1 or Policy 2 scheme with  $2^L$  sources in each group. Under the joint queue size distribution approximation, if Condition (4.74) is satisfied, the steady state distribution of the number of packets transmitted during an epoch, denoted by  $\{\pi_j, 0 \leq j \leq 2^L\}$  is given by the Binomial distribution

$$\pi_j = \binom{2^L}{j} F_{>0}^j F_0^{2^L-j}, \quad 0 \leq j \leq 2^L. \quad (4.88)$$

### Proof

Direct implication of the joint queue size distribution approximation.

Q.E.D.

We are now ready to calculate the performance measures of the systems by applying the steady state distributions given by Theorems 4.3 and 4.4. Henceforth, we assume Condition (4.74) is satisfied.

Recall the three waiting components of a packet defined in Section 4.3. These three waiting components are calculated as follows.

#### The Mean of $W_S^{(1)}$

Since the arrival of a packet is uniformly distributed in a frame, we have

$$E[W_S^{(1)}] = \frac{\alpha-1}{2} . \quad (4.89)$$

#### The Mean of $W^{(2)}$

Let  $W_R$  be the sum of the passive waiting steps of the packets at a source during an epoch at steady state. Then, we have

$$\begin{aligned} E[W_R] &= E[U_{>0}] \sum_{i=2}^{\infty} (i-1) F_i \\ &\quad + F_{>0} \alpha E[A] \sum_{k=1}^{\infty} \frac{k(k-1)}{2} P_{U_{>0}}(k) \\ &\quad + F_0 \alpha E[A] \sum_{k=1}^{\infty} \frac{k(k-1)}{2} P_{U_0}(k) \end{aligned}$$

$$\begin{aligned}
&= E[U_{>0}] \left( \sum_{i=1}^{\infty} i F_i - F_{>0} \right) \\
&+ F_{>0} \alpha E[A] \frac{1}{2} (E[U_{>0}^2] - E[U_{>0}]) \\
&+ F_0 \alpha E[A] \frac{1}{2} (E[U_0^2] - E[U_0]) . \tag{4.90}
\end{aligned}$$

The first term is obtained from the packets already in the source buffer at the beginning of the epoch. The other terms are obtained from the new packets which arrive during the epoch. By incorporating the above result of  $E[W_R]$  and the distribution  $\{\pi_j, 0 \leq j \leq 2^L\}$  given by Theorem 4.4 into the Markov Ratio Limit Theorem, we have

$$E[W^{(2)}] = \frac{2^L E[W_R]}{\sum_{j=0}^{2^L} j \pi_j}, \text{ W.P.1.} \tag{4.91}$$

Since  $\{\pi_j, 0 \leq j \leq 2^L\}$  is a Binomial distribution, we have

$$E[W^{(2)}] = \frac{E[W_R]}{F_{>0}} . \tag{4.92}$$

We skip the algebraic steps and state the mean of  $E[W^{(2)}]$  as follows:

$$\begin{aligned}
E[W^{(2)}] &= \frac{\alpha E[A] E[U_{>0}^2] + \alpha E[U_{>0}]^2 (\text{Var}(A) - E[A])}{2(1 - \alpha E[A] E[U_{>0}])} \\
&+ \frac{E[U_{>0}] \text{Var}(A)}{2E[A]} - \frac{E[U_{>0}]}{2} + \frac{E[U_0^2]}{2E[U_0]} - \frac{1}{2} . \tag{4.93}
\end{aligned}$$

Note that the term  $\sum_{i=1}^{\infty} iF_i$  in Equation (4.90) is calculated by differentiating  $F(z)$  (given by Theorem 4.3) with respect to  $z$  and then evaluate it at  $z = 1$ .

#### The Mean of $W^{(3)}$

By applying the Markov Ratio Limit Theorem, we have

$$E[W^{(3)}] = \frac{\sum_{j=0}^{2^L} E[W_L | j] \pi_j}{\sum_{j=0}^{2^L} j \pi_j}, \text{ W.F.1.} \quad (4.94)$$

#### The Throughput of the Systems

Since each buffer has unlimited buffer capacity, the throughput of a system implemented by a Policy 1 or a Policy 2 scheme is given by

$$S = \lambda = 2^L E[A] . \quad (4.95)$$

#### The Average Delay of the Packets

The average delay of the packets in a system implemented by a Policy 1 scheme is given by

$$E[D] = E[W_S^{(1)}] + (E[W^{(2)}] + E[W^{(3)}])\beta + 1 + R , \quad (4.96)$$

while the average delay of the packets in a system implemented by a Policy 2 scheme is given by

$$E[D] = E[W_S^{(1)}] + (E[W^{(2)}] + E[W^{(3)}])\beta + 1.5 + R , \quad (4.97)$$

where  $E[W_S^{(1)}]$ ,  $E[W^{(2)}]$  and  $E[W^{(3)}]$  are given by Equations (4.89), (4.93) and (4.94), respectively.

### A Fixed-Assignment TDMA Scheme

All along in this chapter, we have assumed that the messages arriving at the sources are single packets. As in Section 4.5, let's now relax this constraint under a Fixed-Assignment TDMA scheme. The messages arriving at the sources are assumed to be composed of multiple packets. The numbers of packets in these messages are also assumed to be independent and identically distributed. Let  $B$  be the number of packets in a message. Set  $P_B(i)$  to be the distribution of  $B$  and  $C_B(z)$  to be the  $z$ -transform of  $P_B(i)$ . Hence, Equations (4.68) - (4.71) become

$$P_{U_0}(k) = \begin{cases} 1, & \text{if } k = 1, \\ 0, & \text{otherwise,} \end{cases} \quad (4.98)$$

$$C_{U_0}(z) = z, \quad (4.99)$$

$$P_{U_{>0}}(k) = P_B(k), \quad (4.100)$$

and

$$C_{U_{>0}}(z) = C_B(z), \quad (4.101)$$

respectively.

By applying the above equations into Theorem 4.3, Condition (4.74) becomes

$$0 < E[A] < \frac{1}{\alpha E[B]}. \quad (4.102)$$

If this condition is satisfied, we have

$$F(z) = \frac{F_0(\phi - \frac{C_B(\phi)}{z})}{1 - \frac{C_B(\phi)}{z}}, \quad (4.103)$$

where

$$F_0 = \frac{1 - \alpha E[A]E[B]}{\alpha E[A] + 1 - \alpha E[A]E[B]}, \quad (4.104)$$

and

$$\phi = C_A(z)^\alpha. \quad (4.105)$$

By applying the above results of  $F(z)$ ,  $\phi$  and  $F_0$  into Corollary 4.1, we have

$$\begin{aligned} E[Q] &= \frac{\alpha^2 E[A]^2 E[B^2] + \alpha E[B](\text{Var}(A) - E[A])}{2(1 - \alpha E[A]E[B])} \\ &\quad + \frac{\alpha E[A]}{2} + \frac{\text{Var}(A)}{2E[A]} - \frac{1}{2}. \end{aligned} \quad (4.106)$$

Since there is only one source in each group, by definitions, we have

$$P_{T_0|j}(k) = \begin{cases} P_{U_0}(k), & \text{if } j = 0, \\ P_{U_{>0}}(k), & \text{if } j = 1, \end{cases} \quad (4.107)$$

and

$$E[W_0|j] = \begin{cases} 0, & \text{if } j = 0, \\ E[B] - 1, & \text{if } j = 1. \end{cases} \quad (4.108)$$

By applying Theorem 4.4, we have

$$\eta_j = \begin{cases} F_0 & , \text{ if } j = 0 , \\ F_{>0} & , \text{ if } j = 1 . \end{cases} \quad (4.109)$$

Let's now investigate the three waiting components of a message.

$E[W_S^{(1)}]$  is the same as given by Equation (4.89).

By applying the results of  $P_{U_0}(k)$  and  $P_{U_{>0}}(k)$  given by Equations (4.98) and (4.100) into Equation (4.93), we have

$$\begin{aligned} E[W^{(2)}] &= \frac{\alpha E[A]E[B^2] + \alpha E[B]^2(\text{Var}(A) - E[A])}{2(1 - \alpha E[A]E[B])} \\ &\quad + \frac{E[B]\text{Var}(A)}{2E[A]} - \frac{E[B]}{2} . \end{aligned} \quad (4.110)$$

In a system implemented by a Fixed-Assignment TDMA scheme, there is no collisions among sources. Hence,

$$E[W^{(3)}] = E[B] - 1 . \quad (4.111)$$

Obviously, the throughput of the system is

$$S = 2^L E[A]E[B] . \quad (4.112)$$

By applying the above results of  $E[W_S^{(1)}]$ ,  $E[W^{(2)}]$  and  $E[W^{(3)}]$  into Equation (4.96), the average delay of the messages is given by

$$\begin{aligned} E[D] &= \frac{\alpha - 1}{2} + \left\{ \frac{\alpha E[A]E[B^2] + \alpha E[B]^2(\text{Var}(A) - E[A])}{2(1 - \alpha E[A]E[B])} \right. \\ &\quad \left. + \frac{E[B]\text{Var}(A)}{2E[A]} - \frac{E[B]}{2} + E[B] - 1 \right\} \alpha + 1 + R \end{aligned}$$

$$= \left\{ \frac{\alpha E[A]E[B]^2 + \alpha E[B]^2(\text{Var}(A) - E[A])}{2(1-\alpha E[A]E[B])} + \frac{E[B]\text{Var}(A)}{2E[A]} + \frac{E[B]}{2} \right\} \alpha - \frac{\alpha-1}{2} + R \quad (4.113)$$

Note that under a Fixed-Assignment TDMA scheme, the construction of superframes is not necessary, therefore,  $\alpha = \beta$ .

### Numerical Results

To demonstrate the performance of the system, we again consider a numerical example of 64 sources as in Section 4.5. However, the arrival process of packets at a source is now assumed to be Poisson,

$$P_A(i) = \frac{\lambda_s^i e^{-\lambda_s}}{i!}.$$

When Policy 1 is implemented, Figure 4.19 plots the average queue size,  $E[Q]$ , at each source versus the throughput,  $S$ , of the system. Figures 4.20 and 4.21 plot the average delay of the packets versus the throughput,  $S$ , of the system when  $R = 0$  and  $R = 12$ , respectively.

When the system is implemented by Policy 2, a similar set of curves are plotted in Figures 4.22-4.24.

The characteristics of these two sets of curves are shown to be similar to the characteristics of those in Section 4.5. However, the two sets of delay-throughput curves obtained in this section are much higher than those in Section 4.5 (especially when the throughput,  $S$ , is high) because in this section, all the arriving packets at the sources are accepted while in Section 4.5, only the first arriving packet at each



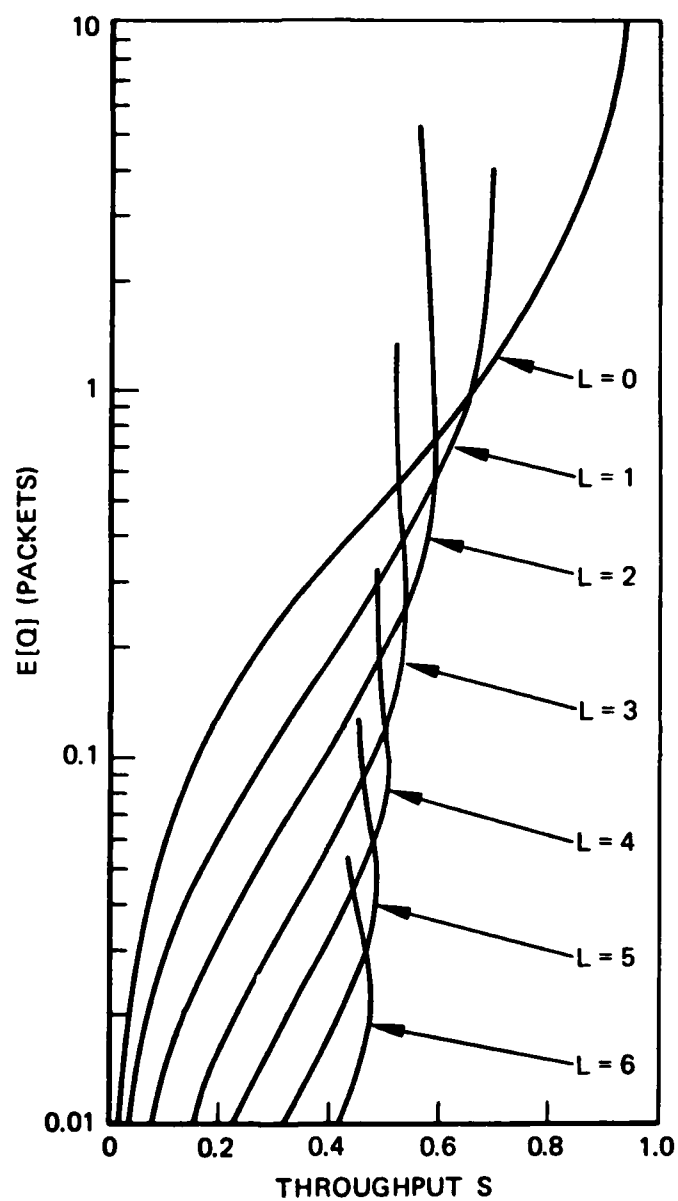


Figure 4.19.  $E[Q]$  vs Throughput Curves under Policy 1.

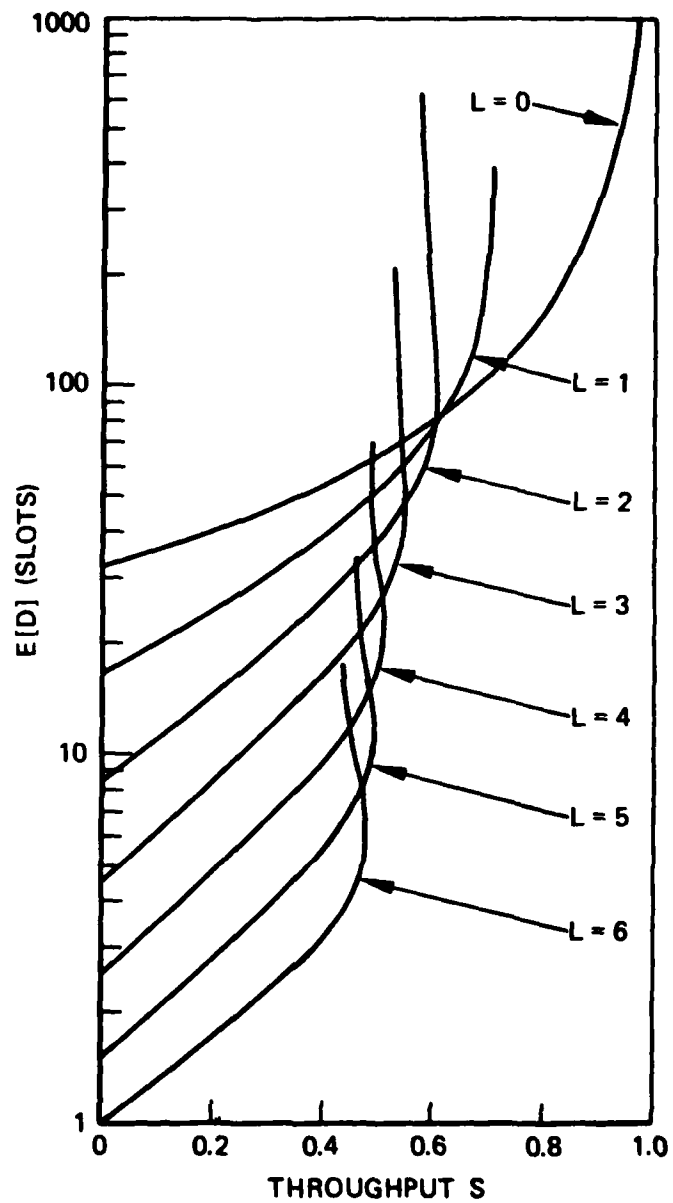


Figure 4.20. Delay - Throughput Curves under Policy 1,  $R = 0$ .

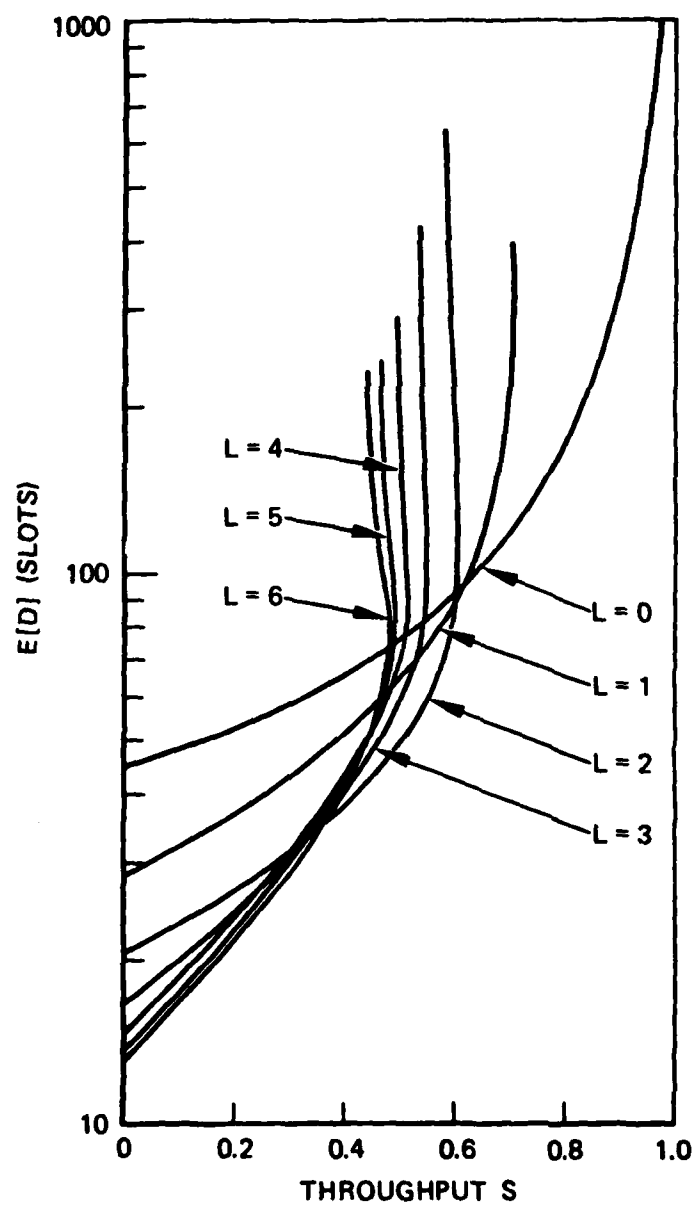


Figure 4.21. Delay - Throughput Curves under Policy 1,  $R = 12$ .

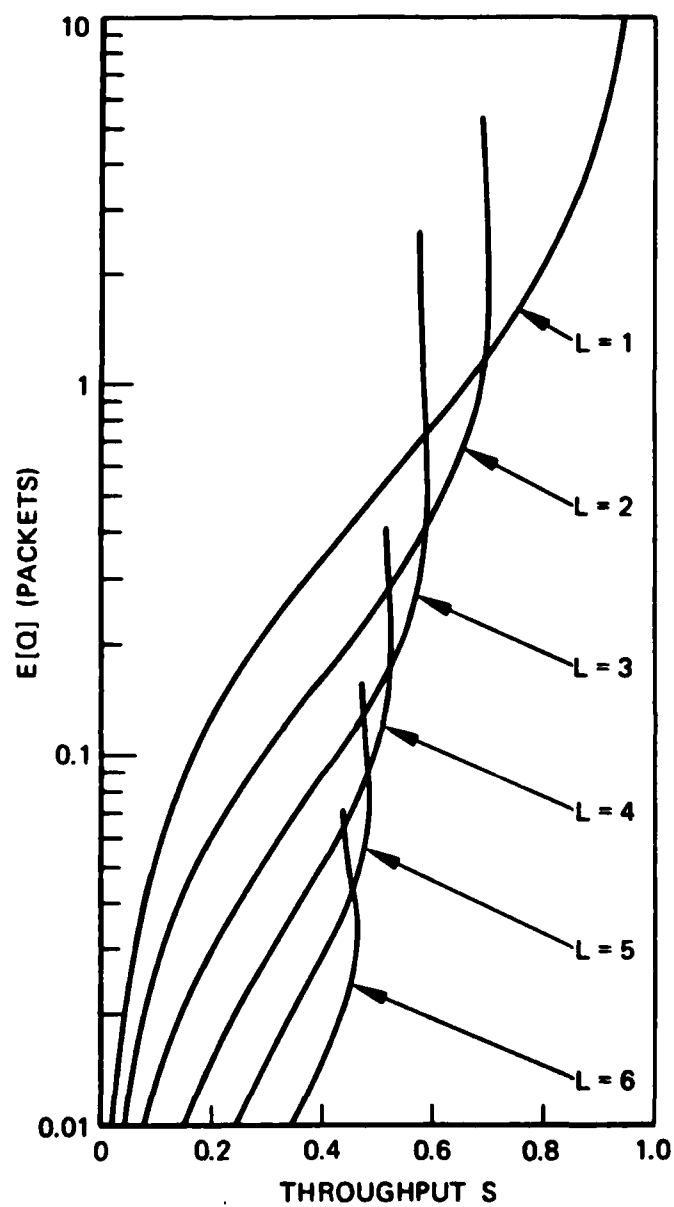


Figure 4.22.  $E[Q]$  vs Throughput Curves under Policy 2.

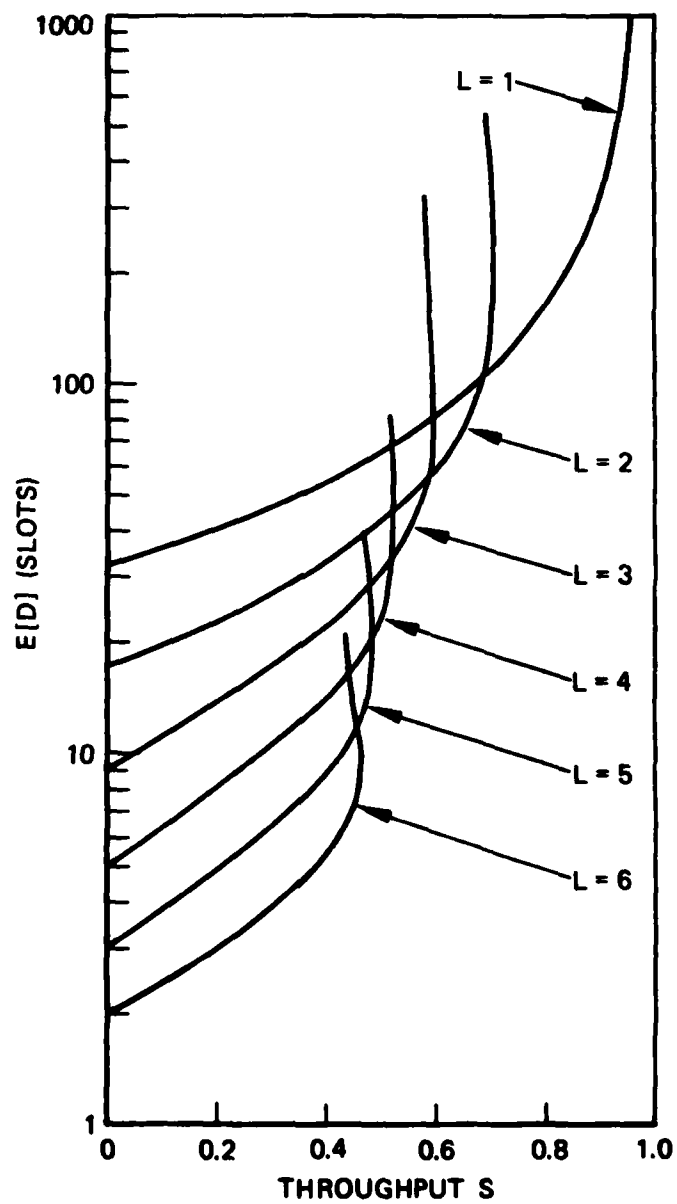


Figure 4.23. Delay – Throughput Curves under Policy 2,  $R = 0$ .

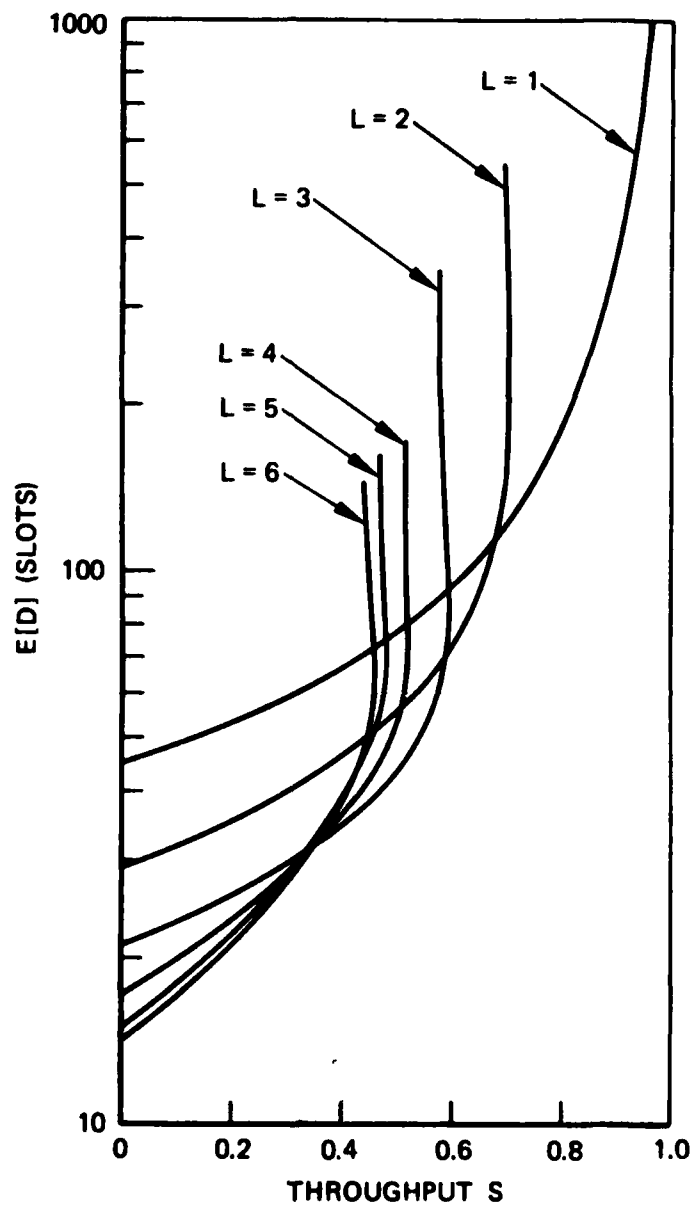


Figure 4.24. Delay - Throughput Curves under Policy 2,  $R = 12$ .

source during an on-going epoch is accepted.

#### Justification of the Joint Queue Size Distribution Approximation

When we investigated the waiting time of a packet, we split it into three components,  $W_S^{(1)}$ ,  $W_S^{(2)}$  and  $W_S^{(3)}$ . The approximation is applied only to obtain the mean of  $W_S^{(2)} + W_S^{(3)}$ . We obtained the mean of  $W_S^{(1)}$  by exact analysis. Hence, we only need to justify the obtained mean of  $W_S^{(2)} + W_S^{(3)}$ .

From Equation (4.95), we observe that as the throughput,  $S$ , approaches 0, the average input rate also approaches 0. If we let  $E[A]$  approach 0, in Equations (4.93) and (4.94), the values of  $E[W^{(2)}]$  and  $E[W^{(3)}]$ , will both go to 0. Hence, the approximation is exact when the throughput,  $S$ , is 0. It should also give good results in the neighborhood of 0 because both  $E[W^{(2)}]$  and  $E[W^{(3)}]$  are continuous, smooth functions of  $S$ . In Figures 4.25 and 4.26, simulation and approximation results are plotted for the systems implemented by Policy 2 with  $L = 2$  and  $L = 3$ , respectively. The approximation results are observed to be very close to simulation results (virtually identical) if  $S < 0.6$ . If  $S \geq 0.6$ , the approximation is moderate. However, in the latter range of  $S$ , the Fixed-Assignment TDMA scheme (exact analysis and no approximation is required) yields generally a better delay-throughput performance curve than all the other schemes in each family.

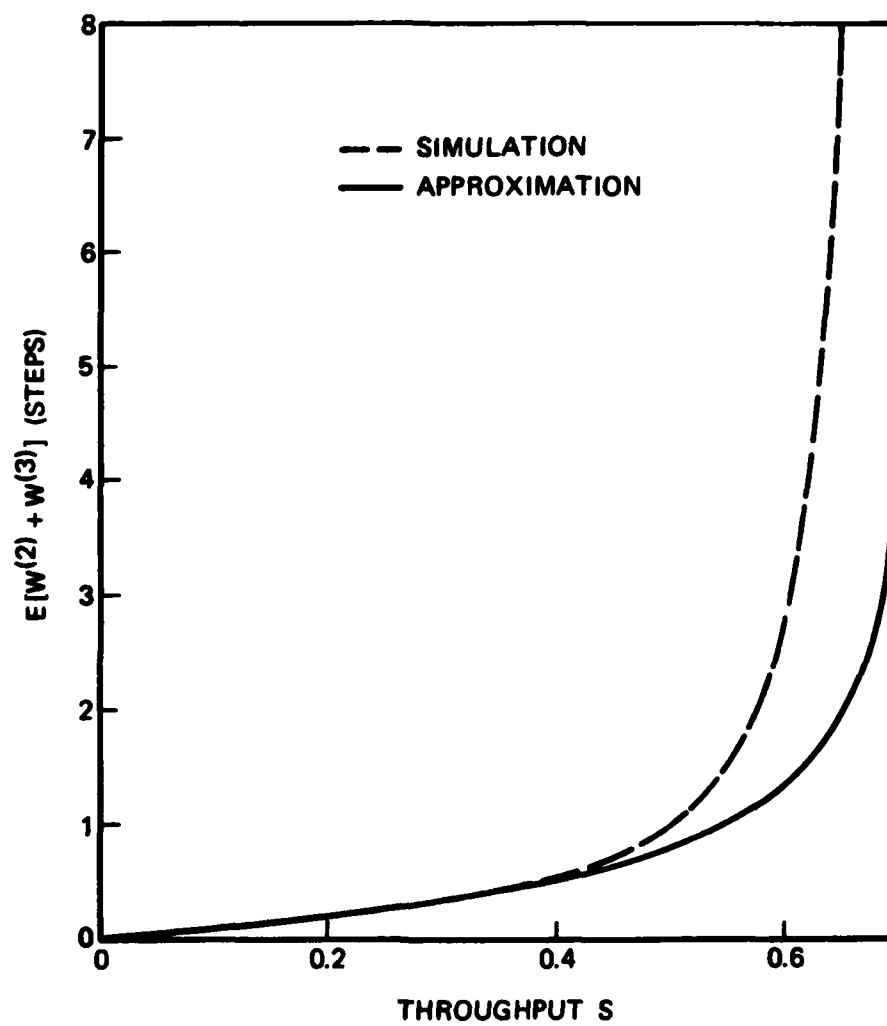


Figure 4.25.  $E[W^{(2)} + W^{(3)}]$  vs Throughput Curves under Policy 2 by Simulation and Approximation,  $L = 2$ .



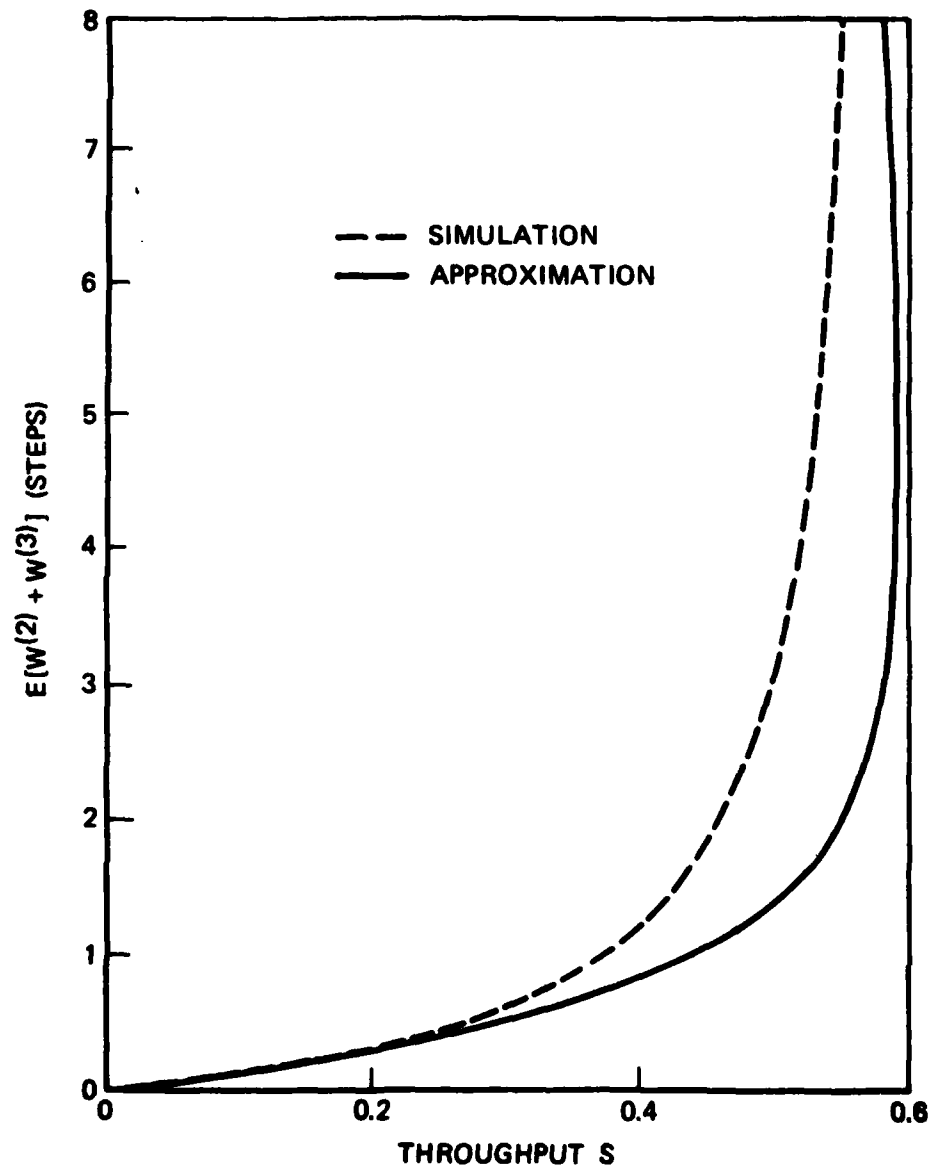


Figure 4.26.  $E[W(2) + W(3)]$  vs Throughput Curves under Policy 2 by Simulation and Approximation,  $L = 3$ .

## CHAPTER V

### FIXED RESERVATION SCHEMES WITH PREEMPTIVE PRIORITIES

#### 5.1 Introduction

In the last chapter, we discussed a class of hybrid access-control schemes used to control the sharing of a broadcast communication channel by a number of sources. Messages arriving at the sources are assumed to be single packets. If messages arriving at the sources are composed of multiple packets, then this class of hybrid access-control schemes may not be efficient because of the following two reasons. First, the packets of each message have to be transmitted independently over the channel. The delay of each message is then the sum of the delays of all the packets of the message, therefore, could be too long to be acceptable. Second, the packets of each message are transmitted independently over the channel, therefore, packet reassembly is necessary at the destination. Consequently, to share a channel efficiently on a store-and-forward basis, we consider reservation schemes.

If the message traffic is steady, Fixed-Assignment TDMA schemes or Polling schemes are efficient. These schemes are actually predetermined reservation schemes. Under these schemes, sources are permitted to transmit messages in a fixed cyclic order and there is no need to send reservation information. However, if the message traffic is bursty (statistically fluctuating), then these schemes may not be suitable because of low utilization of the channel. (Both schemes were briefly discussed in Chapter I.) Subsequently, we require sources

which have messages to transmit send reservation information to reserve future slots.

The reservation schemes we are going to study are described as follows. The channel time frame is divided into two portions, one portion for making reservations by sources which have messages to transmit and the other portion for actual message transmission. The reservation portion is a broadcast channel, so that by listening to it, every source can obtain information about the messages which have successfully made reservations. Based on the successfully received reservation information sent by all the sources, transmission is scheduled by following a previously agreed policy.

The reservation traffic in the reservation portion of the channel can be controlled by one of the access-control schemes discussed in Chapter IV. In this chapter, we are interested in studying the traffic in the other portion of the channel.

In Section 5.2, we will introduce the system model and investigate the delay of a message intuitively. In Section 5.3, we will analyze the average delay of the messages. In Section 5.4, we will introduce a preemptive priority discipline to allocate future slots for messages which have successfully sent reservation information. The average delays of messages with different priorities are studied. A numerical example is presented.

## 5.2 The System Models

In the following, we describe the channel model which will be used in this chapter. The protocol of the reservation schemes and

the traffic model are presented. Also, the delay of a message is investigated intuitively.

#### The Channel Model

The channel is time-slotted, that is, the channel time is divided into equal durations called slots. These slots are designed long enough so that exactly one message packet can be transmitted in one slot. The slots are arranged to form consecutive blocks of slots called frames. Each frame contains the same number,  $\alpha$ , of slots.

In every frame, a fixed portion,  $\gamma$  consecutive slots, is used for reservations. Sources that have messages to transmit make reservations in these reservation periods. These reservation periods serve as a broadcast channel so that every source can listen to it and obtain information about the messages which have successfully made reservations. The traffic in the reservation periods can be controlled by the access-control schemes discussed in the last chapter. The size of the reservation packets may of course be much smaller than the size of the message packets. Hence, each slot in the reservation periods may be subdivided into mini slots such that each mini slot is long enough for the transmission of one reservation packet.

The other portion of every frame,  $\beta$ ,  $\beta = \alpha - \gamma$ , consecutive slots is used for actual message transmission. These periods are used by messages which have successfully made reservations. Messages which have successfully made reservations in different reservation periods are transmitted in a first come first served order. Messages which have made reservations in the same period are to be transmitted orderly according

to a previously agreed policy. Hence, there is no contention in these service periods.

Set  $R$  (slots) to be the propagation delay of the channel. Hence,  $R$  slots after the transmission of a reservation period, this reservation period is completely received by every source. Depending on the construction of the channel frame, a reservation period is received during a service period or during another reservation period. We set  $M$  (slots) to be the number of remaining message slots in the service period right after a reservation period has been completely received and  $R_R$  (slots) to be the interval from the end of a reservation period to the first message slot right after the reservation period has been completely received.

Consider the first case. A reservation period is completely received by every source during a service period. (This is illustrated by Case 1 in Figure 5.1.)  $R_R = R$  and  $M \leq \beta$ .

Consider the second case. A reservation period is completely received by every source during another reservation period. (This is illustrated by Case 2 in Figure 5.1.)  $M = \beta$  and  $R + \gamma \geq R_R \geq R$ .

#### Protocol and the Traffic Model

When a message has arrived at a source, the source will start to make reservation for the message in the upcoming reservation period. Reservation has to be made by following an access-control scheme established for the broadcast reservation periods of the channel. When the reservation has been successfully made and acknowledged by

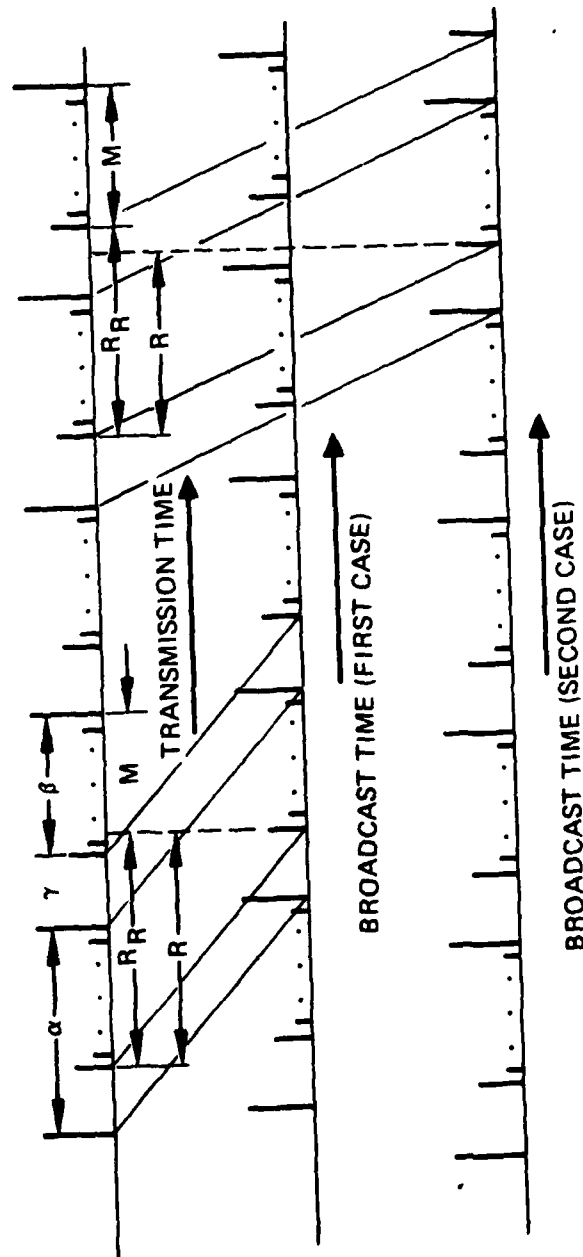


Figure 5.1. An Example illustrating the Effect of Propagation Delay on the Broadcast of a Broadcast Channel under a Fixed Reservation Scheme. In the First Case,  $R_R = R$  and  $M < \beta$ . In the Second Case,  $R + \gamma \geq R_R \geq R$  and  $M = \beta$ .

broadcast, future slots are reserved for the message according to a previously agreed policy. Hence, messages which have successfully made reservations form a channel queue.

#### The Delay of a Message

The delay of a message, denoted by  $D$  (slots), is measured from the start of the first slot after its arrival to the time it is completely received. This can be split into two parts.

$$D = D^{(R)} + D^{(S)} . \quad (5.1)$$

The first part, denoted by  $D^{(R)}$  (slots), is the reservation delay.

The second part, denoted by  $D^{(S)}$  (slots), is the service delay.

The reservation delay is measured from the start of the first slot after the arrival of the message to the start of the first message slot after the broadcast of the reservation period in which the reservation of the message is successfully made.  $D^{(R)}$  can be further decomposed into

$$D^{(R)} = \hat{U} + \hat{D} + \gamma + R_R . \quad (5.2)$$

The first term,  $\hat{U}$  (slots), is measured from the slot after the arrival of the message to the start of the first upcoming reservation period. The distribution of  $\hat{U}$  depends on the statistics of the arrivals of messages at the sources. If the arrival of a new message is uniformly distributed over any frame, then the distribution of  $\hat{U}$  is

$$P_{\hat{U}}(i) = \begin{cases} 1/\alpha & , \quad 0 \leq i \leq \alpha - 1, \\ 0 & , \quad \text{otherwise,} \end{cases} \quad (5.3)$$

and

$$E[\hat{U}] = (\alpha - 1)/2 \quad (5.4)$$

The second term,  $\hat{D}$  (slots), is the reservation contention delay. It is measured from the start of the first upcoming reservation period to the start of the reservation period in which the reservation is successfully made. If there is no reservation contention,  $\hat{D} = 0$ .

The last two terms are the duration of the reservation period in which the reservation is successfully made, its propagation delay and the time elapsed before the first upcoming message slot.

The service delay, denoted by  $D^{(S)}$  (slots), in Equation (5.1) begins right after the period  $D^{(R)}$  and ends when the message is completely received. It can be decomposed into

$$D^{(S)} = W^{(L)} + W^{(G)} + T \quad (5.5)$$

Observe that in each reservation period, reservations are made for a group of messages. Hence, we have message group arrivals at the channel queue. In each group, the message which is transmitted first is called the leader of the group. Recall that messages which have successfully made reservations in different reservation periods are transmitted in a first come first served order. Hence, in a group, the leader's service waiting time is the channel queue immediately before the group's reservation period has been completely received. This is the first term of  $D^{(S)}$ .

The second term corresponds to those messages in the same group but are transmitted first.



The last term is the transmission of the message and the channel's propagation delay.

We are going to study the delays of the messages after their reservations have been successfully made. We make the following assumptions.

#### Assumption 5.1

The numbers of messages which have successfully made reservations in a sequence of reservation periods are independent and identically distributed. Let  $A$  be the number of messages which have successfully made reservations in a reservation period. The distribution of  $A$  is denoted by  $P_A(i)$  and the  $z$ -transform of  $P_A(i)$  is denoted by  $C_A(z)$ .

#### Assumption 5.2

The numbers of packets in messages are independent and identically distributed. Let  $B$  be the number of packets in a message. The distribution of  $B$  is denoted by  $P_B(i)$  and the  $z$ -transform of  $P_B(i)$  is denoted by  $C_B(z)$ .

### 5.3 Bounds on the Average Message Delay

In the last section, we split the total delay of a message into two parts, the reservation delay and the service delay. The reservation delay is described by Equations (5.2) - (5.4). Let's now investigate the service delay expressed by Equation (5.5).

In the measurement of all the delay terms,  $D^{(S)}$ ,  $W^{(L)}$ ,  $W^{(G)}$  and  $T$ , both types of slots (reservation slots and service slots) are counted. The technique employed in our analysis requires that in the measurement of all the relevant delay terms, only service slots are

counted. Hence, we denote each relevant delay term, in which only service slots are counted, by a tilde in our analysis. When the analysis is finished, we will convert all the useful results back to terms in which both types of slots are counted. From  $W^{(L)}$  and  $W^{(G)}$ , we have  $\tilde{W}^{(L)}$  and  $\tilde{W}^{(G)}$  when only service slots are counted. From Equation (5.5), we define

$$\tilde{D}^{(S)} = \tilde{W}^{(L)} + \tilde{W}^{(G)} + B. \quad (5.6)$$

Hence,  $\tilde{D}^{(S)}$  is the service delay less the propagation delay of a message when only service slots are counted. Before we analyze  $\tilde{W}^{(L)}$  and  $\tilde{W}^{(G)}$  of Equation (5.6), we make the following definitions:

$G$  - the number of packets which have successfully made reservations in a reservation period. From Assumptions 5.1 and 5.2, the numbers of packets which have successfully made reservations in a sequence of reservation periods are independent and identically distributed. Set  $P_G(i)$  to be the distribution of  $G$  and  $C_G(z)$  to be the  $z$ -transform of  $P_G(i)$ . Then we have

$$P_G(i) = \sum_{k=0}^{\infty} P_B^{(k)}(i) P_A(k), \quad (5.7)$$

where  $P_B^{(k)}(i)$  is the  $k$ -time convolution of  $P_B(i)$ . By definition,

$$\begin{aligned} C_G(z) &= \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} P_B^{(k)}(i) P_A(k) z^i \\ &= \sum_{i=0}^{\infty} P_A(k) (C_B(z))^k \\ &= C_A(C_B(z)), \quad |z| \leq 1. \end{aligned} \quad (5.8)$$

Hence,

$$E[G] = E[A]E[B] . \quad (5.9)$$

$\rho$  - the throughput of the service portion of the channel. This is equal to  $E[G]/\beta$ .

$\tilde{A}$  - the number of messages in a group. From Assumptions 5.1 and 5.2, the numbers of messages in a sequence of groups are independent and identically distributed. Set  $P_{\tilde{A}}(i)$  to be the distribution of  $\tilde{A}$  and  $C_{\tilde{A}}(z)$  to be the z-transform of  $P_{\tilde{A}}(i)$ . From Assumption 5.1, we have

$$P_{\tilde{A}}(i) = \begin{cases} P_A(i)/(1 - P_A(0)), & \text{if } 0 < i < \infty, \\ 0, & \text{otherwise,} \end{cases} \quad (5.10)$$

$$E[\tilde{A}] = E[A]/(1 - P_A(0)) , \quad (5.11)$$

and

$$C_{\tilde{A}}(z) = (C_A(z) - P_A(0))/(1 - P_A(0)) . \quad (5.12)$$

$\tilde{G}$  - the number of packets in a group. From Assumptions 5.1 and 5.2, the numbers of packets in a sequence of message groups are independent and identically distributed. Set  $P_{\tilde{G}}(i)$  to be the distribution of  $\tilde{G}$  and  $C_{\tilde{G}}(z)$  to be the z-transform of  $P_{\tilde{G}}(i)$ . Then we have

$$P_{\tilde{G}}(i) = \sum_{k=1}^{\infty} P_B^{(k)}(i) P_{\tilde{A}}(k) , \quad (5.13)$$

and

$$E[\tilde{G}] = E[\tilde{A}] E[B] . \quad (5.14)$$

$\tau$  - the number of reservation periods between two successive groups of messages. From Assumption 5.1, the numbers of reservation periods between successive groups of messages are independent and identically distributed. Let  $P_\tau(i)$  be the distribution of  $\tau$ .

We have

$$P_\tau(i) = \begin{cases} P_A(0)^{i-1} (1 - P_A(0)), & \text{if } 0 < i < \infty, \\ 0, & \text{otherwise,} \end{cases} \quad (5.15)$$

and

$$E[\tau] = 1/(1 - P_A(0)) . \quad (5.16)$$

#### Bounds on the Mean of $W^{(L)}$

Define  $\tilde{W}_n^{(L)}$  to be the waiting time, counted in service slots, of the leader in the  $n^{\text{th}}$  group. Also define  $\tilde{G}_n$  to be the number of packets in the  $n^{\text{th}}$  group and  $\tau_n$  to be the number of reservation periods between the  $n^{\text{th}}$  and  $n+1^{\text{th}}$  groups. For the Markov Chain  $\{\tilde{W}_n^{(L)}, n > 0\}$ , we have

$$\tilde{W}_{n+1}^{(L)} = [\tilde{W}_n^{(L)} + \tilde{G}_n - \tau_n \beta]^+ , \quad (5.17)$$

where

$$[x]^+ = \text{Max}(0, x) .$$

#### Theorem 5.1

Under a Fixed Reservation scheme, if  $\rho < 1$ , a limiting distribution of  $\tilde{W}_n^{(L)}$  exists. Its z-transform, denoted by  $L(z)$ , is given by

$$L(z) = \frac{\beta(1-\rho)(1-z)}{C_G(z) - z^\beta} \prod_{r=1}^{\beta-1} \frac{z - \eta_r}{1 - \eta_r} , \quad (5.18)$$

and  $\eta_r$ ,  $r = 1, 2, \dots, \beta-1$ , are the distinct  $\beta-1$  roots with  $|\eta_r| < 1$ , obtained by the limit

$$\eta_r = \lim_{\omega \rightarrow 1^-} \eta_r(\omega), \quad (5.19)$$

where  $\eta_r(\omega)$ ,  $r = 1, 2, \dots, \beta$ , are the  $\beta$  distinct roots of the functional equation

$$z^\beta = \omega C_G(z). \quad (5.20)$$

Furthermore,

$$E[\tilde{W}^{(L)}] = \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] + \frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1+\eta_r}{1-\eta_r}, \quad (5.21)$$

where  $\text{Var}(G)$  is the variance of  $G$ .

#### Proof

The recurrence Relationship (5.17) is identical to that of the waiting times of the customers in a GI/G/1 queueing system with customers interarrival times  $\beta\tau_n$ ,  $n > 0$ , and service times  $\tilde{G}_n$ ,  $n > 0$ . By applying Lindley's Theorem to this GI/G/1 queueing system, we conclude that the Markov Chain  $\{\tilde{W}_n^{(L)}, n > 0\}$  has a proper limiting distribution, independent of  $\tilde{W}_1^{(L)}$ , if and only if

$$E[\tilde{G}]/E[\tau]\beta < 1. \quad (5.22)$$

Apply Equations (5.11), (5.14) and (5.16) into the left side of Inequality (5.22), we have

$$E[\tilde{G}]/E[\tau]\beta = E[A]E[B]/\beta. \quad (5.23)$$

Incorporating Equation (5.9) into Equation (5.23), we have

$$E[\tilde{G}]/E[\tau]\beta = E[G]/\beta . \quad (5.24)$$

Apply Equation (5.24) to the left side of Inequality (5.22) and by definition, we have

$$\rho = E[G]/\beta < 1 .$$

Hence, if  $\rho < 1$ , the GI/G/1 queueing system has a steady state distribution and so does the Markov Chain  $\{\tilde{W}_n^{(L)}, n > 0\}$ .

Now, let's set  $X_n$  to be the channel queue immediately before the  $n^{\text{th}}$  reservation period is completely received. The Markov Chain  $\{X_n, n > 0\}$  is governed by the recurrence relationship

$$X_{n+1} = [X_n + G_n - \beta]^+ . \quad (5.25)$$

Observe that  $\{\tilde{W}_n^{(L)}, n > 0\}$  is obtained by sampling  $\{X_n, n > 0\}$  at those reservation periods which have group arrivals. However, these group arrivals are described by a discrete-time renewal point process with interarrival times  $\tau_n, n > 0$ , which are each geometrically distributed as described by Equation (5.15). This constitutes a memoryless occurrence of a random sequence of events. Hence,  $X_n$  and  $\tilde{W}_n^{(L)}$  have identical steady state distribution. It can be shown (see [24]) the z-transform of the steady state distribution of  $X_n$ , denoted by  $C_X(z)$ , in recurrence Relationship (5.25) is independent of  $X_1$  and  $n$ , and is given by

$$C_X(z) = \frac{\beta(1-\rho)(1-z)}{C_G(z) - z^\beta} \prod_{r=1}^{\beta-1} \frac{z - \eta_r}{1 - \eta_r} , \quad (5.26)$$

and  $\eta_r, r = 1, 2, \dots, \beta-1$ , are the distinct  $\beta-1$  roots with  $|\eta_r| < 1$ , obtained by the limit

$$\eta_r = \lim_{\omega \rightarrow 1} \eta_r(\omega),$$

where  $\eta_r(\omega)$ ,  $r = 1, 2, \dots, \beta$ , are the distinct roots of the functional equation

$$z^\beta = \omega C_G(z).$$

The mean of  $\tilde{W}_G^{(L)}$  in Equation (5.21) is obtained by differentiating the right side of Equation (5.18) with respect to  $z$ , and then evaluate it at  $z=1$ . Q.E.D.

### Theorem 5.2

If  $\rho < 1$ , the Markov Chain  $\{\tilde{W}_n^{(L)}, n > 0\}$  has a steady state mean  $E[W^{(L)}]$ , which is upperbounded by

$$E[\tilde{W}^{(L)}]_u \stackrel{\Delta}{=} \text{Var}(G)/2\beta(1-\rho), \quad (5.27)$$

and lower bounded by

$$E[W^{(L)}]_l \stackrel{\Delta}{=} \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G]. \quad (5.28)$$

### Proof

From Theorem 5.1, the Markov Chains  $\{X_n, n > 0\}$  and  $\{\tilde{W}_n^{(L)}, n > 0\}$  in recurrence Relationships (5.25) and (5.17), respectively, have been shown to have identical steady state distributions when  $\rho < 1$ . Hence, they have the same steady state mean. An upper bound of  $E[x]$ , denoted by  $E[x]_u$ , is also an upper bound of  $E[W^{(L)}]$ . By applying Kingman's result (see [32]), we have

$$E[x] \leq E[x]_u = \text{Var}(G)/2\beta(1-\rho). \quad (5.29)$$

Hence,

$$E[\tilde{W}^{(L)}] \leq E[\tilde{W}^{(L)}]_u = \text{Var}(G)/2\beta(1-\rho) . \quad (5.30)$$

From Equation (5.21), we have

$$E[\tilde{W}^{(L)}] = \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] + \frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1 + \eta_r}{1 - \eta_r} . \quad (5.31)$$

Observe the last term of the above equation. Set

$$\eta_r = a_r + jb_r, \quad \text{for } 1 \leq r \leq \beta - 1.$$

We have

$$\frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1 + \eta_r}{1 - \eta_r} = \frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1 - (a_r^2 + b_r^2) + 2jb_r}{(1 - a_r)^2 + b_r^2} . \quad (5.32)$$

Since  $|\eta_r| < 1$ , we have

$$a_r^2 + b_r^2 < 1, \quad \text{for } 1 \leq r \leq \beta - 1. \quad (5.33)$$

By applying Inequality (5.33) into Equation (5.32), we have

$$\frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1 + \eta_r}{1 - \eta_r} > 0. \quad (5.34)$$

By applying Inequality (5.34) into Equation (5.31), we have

$$E[\tilde{W}^{(L)}] > \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] . \quad (5.35)$$

If  $\rho = 0$ , both sides of Inequality (5.35) vanish. Hence,

$$E[\tilde{W}^{(L)}] \geq E[\tilde{W}^{(L)}]_l = \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] .$$

Q.E.D.

Note that

$$E[\tilde{W}^{(L)}]_u - E[\tilde{W}^{(L)}]_l = E[G]/2 . \quad (5.36)$$



Also note that from Equation (5.31),

$$E[\tilde{W}^{(L)}] = E[\tilde{W}^{(L)}]_{\lambda}, \text{ if } \beta = 1. \quad (5.37)$$

#### The Mean of $W^{(G)}$

We now consider the group waiting time of a message. We have the following theorem.

#### Theorem 5.3

Under a Fixed Reservation scheme, the average group waiting time of a message (induced by the other messages in the same group), is given by

$$E[\tilde{W}^{(G)}] = \frac{1}{2} E[B] \left( \frac{E[A^2]}{E[A]} - 1 \right). \quad (5.38)$$

#### Proof

Let  $\tilde{A}_j$  be the number of messages in the  $j^{\text{th}}$  group. The random variables  $\tilde{A}_j$ ,  $j > 0$ , are independent and identically distributed with distribution  $P_{\tilde{A}}(i)$  given by Equation (5.10). Consider the sum

$$S_0 \stackrel{\Delta}{=} 0,$$

$$S_n \stackrel{\Delta}{=} \sum_{j=1}^n \tilde{A}_j, \text{ if } n > 0.$$

If we take the messages to be the time index  $t$ , then the stochastic process  $\{v_t, t \in [1, \infty)\}$  with

$$v_t \stackrel{\Delta}{=} \text{Max}\{n: S_n < t\},$$

$$v_1 \stackrel{\Delta}{=} 0,$$

is a renewal process. The stochastic variable

$$\xi_t = t - v_t - 1,$$

the past life time at  $t$ , is the number of messages in the group transmitted before  $t$ . Denote the limit of  $\xi_t$  by  $\xi$  and its distribution by  $P_\xi(i)$ . From Renewal theory (see [33]), we have

$$P_\xi(i) = \frac{1}{\mu} \sum_{j=i+1}^{\infty} P_{\tilde{A}}(j), \quad (5.39)$$

where

$$1/\mu = 1/E[\tilde{A}]. \quad (5.40)$$

The  $z$ -transform of  $P_\xi(i)$ , denoted by  $C_\xi(z)$  is

$$\begin{aligned} C_\xi(z) &= \frac{1}{\mu} \sum_{i=0}^{\infty} z^i \sum_{j=i+1}^{\infty} P_{\tilde{A}}(j) \\ &= (1 - C_{\tilde{A}}(z))/E[\tilde{A}](1-z) \\ &= (1 - C_A(z))/E[A](1-z). \end{aligned} \quad (5.41)$$

The waiting time induced by the messages in the same group of  $t$ , denoted by  $\tilde{W}_t^{(G)}$ , is the transmission time of those messages in the same group of  $t$  but transmitted before  $t$ . Set  $H(z)$  to be the  $z$ -transform of the distribution of  $\tilde{W}_t^{(G)}$  at the limit  $t \rightarrow \infty$ , we have

$$\begin{aligned} H(z) &= \sum_{j=0}^{\infty} P_\xi(j) C_B(z)^j \\ &= (1 - C_{\tilde{A}}(C_B(z)))/E[\tilde{A}](1 - C_B(z)) \\ &= (1 - C_A(C_B(z)))/E[A](1 - C_B(z)) \end{aligned} \quad (5.42)$$

By differentiating  $H(z)$  with respect to  $z$  and then evaluate it at  $z=1$ , we have

$$E[\tilde{W}^{(G)}] = \frac{1}{2} E[B] \left( \frac{E[A^2]}{E[A]} - 1 \right).$$

Q.E.D.

#### The Mean of $\tilde{D}^{(S)}$

From Equation (5.6), we have

$$E[\tilde{D}^{(S)}] = E[\tilde{W}^{(L)}] + E[\tilde{W}^{(G)}] + E[B]. \quad (5.43)$$

By applying Theorem 5.3 to the second term, we have

$$E[\tilde{D}^{(S)}] = E[\tilde{W}^{(L)}] + \frac{1}{2} E[B] \left( \frac{E[A^2]}{E[A]} + 1 \right) \quad (5.44)$$

#### Bounds on the Mean of $D^{(S)}$

Next, we introduce the following theorem which will be used to convert  $E[\tilde{D}^{(S)}]$  (in which only service slots are counted) back to  $E[D^{(S)}]$  (in which both types of slots are counted).

#### Theorem 5.4

Let  $\phi(i,j)$  be a non-decreasing function with respect to both  $i$  and  $j$ . Also let  $P_\Gamma(i)$  be the distribution of a random variable  $\Gamma$  over the space  $\{i: i > 0\}$ . Define

$$\psi = \sum_{j=0}^{\Delta} \sum_{i=j\delta+\Delta+1}^{(j+1)\delta+\Delta} \phi(i,j) P_\Gamma(i), \text{ for some } \delta > 0 \quad (5.45)$$

and  $\Delta \geq 0$ .

Then

$$\psi \leq \psi_u = \sum_{i=\Delta+1}^{\Delta} \phi\left(i, \frac{i-\Delta-1}{\delta}\right) P_\Gamma(i), \quad (5.46)$$

and

$$\psi \geq \psi_{\ell} = \sum_{i=\Delta+1}^{\Delta} \phi\left(i, \frac{i-\Delta}{\delta} - 1\right) P_{\Gamma}(i) . \quad (5.47)$$

Furthermore, if  $\delta = 1$ ,

$$\psi_{\ell} = \psi = \psi_u . \quad (5.48)$$

### Proof

Observe the interval

$$j\delta + \Delta + 1 \leq i \leq (j+1)\delta + \Delta .$$

We have

$$j \leq \frac{i-\Delta-1}{\delta} , \quad (5.49)$$

and

$$\frac{i-\Delta}{\delta} - 1 \leq j . \quad (5.50)$$

Since  $\phi(i,j)$  is non-decreasing with respect to both  $i$  and  $j$ , by applying Relationship (5.49) into (5.45), we have

$$\begin{aligned} \psi \leq \psi_u &= \sum_{j=0}^{\infty} \sum_{i=j\delta+\Delta+1}^{(j+1)\delta+\Delta} \phi\left(i, \frac{i-\Delta-1}{\delta}\right) P_{\Gamma}(i) \\ &= \sum_{i=\Delta+1}^{\infty} \phi\left(i, \frac{i-\Delta-1}{\delta}\right) P_{\Gamma}(i) . \end{aligned}$$

By applying Relationship (5.50) into (5.45), we have

$$\begin{aligned} \psi \geq \psi_{\ell} &= \sum_{j=0}^{\infty} \sum_{i=j\delta+\Delta+1}^{(j+1)\delta+\Delta} \phi\left(i, \frac{i-\Delta}{\delta} - 1\right) P_{\Gamma}(i) \\ &= \sum_{i=\Delta+1}^{\infty} \phi\left(i, \frac{i-\Delta}{\delta} - 1\right) P_{\Gamma}(i) . \end{aligned}$$

Obviously, if  $\delta = 1$ ,  $\psi_\ell = \psi = \psi_u$ .

Q.E.D.

Now, we convert  $E[\tilde{D}^{(S)}]$  back to  $E[D^{(S)}]$  by the following theorem.

Theorem 5.5

Under a Fixed Reservation scheme, if  $\rho < 1$ , the mean of  $D^{(S)}$  is upper bounded by

$$E[D^{(S)}]_u \stackrel{\Delta}{=} \frac{\alpha}{\beta} E[\tilde{D}^{(S)}] + [(\beta-1) - M] \left( \frac{\alpha}{\beta} - 1 \right) + R, \quad (5.51)$$

and lower bounded by

$$E[D^{(S)}]_\ell \stackrel{\Delta}{=} \frac{\alpha}{\beta} E[\tilde{D}^{(S)}] - M \left( \frac{\alpha}{\beta} - 1 \right) + R. \quad (5.52)$$

Proof

From Theorem 5.2, if  $\rho < 1$ , the Markov Chain  $\{\tilde{w}_n^{(L)}, n > 0\}$  has a steady state distribution. Hence,  $\tilde{D}^{(S)}$  has a steady state distribution. Denote the steady state distribution of  $\tilde{D}^{(S)}$  by  $P(i)$ , we have

$$E[D^{(S)}] = \sum_{i=1}^M i P(i) + \sum_{j=0}^{\infty} \sum_{i=j\beta+M+1}^{(j+1)\beta+M} (i+(j+1)\gamma) P(i) + R. \quad (5.53)$$

By applying Relationship (5.46) of Theorem 5.4 to the second term of Equation (5.53), we have

$$\begin{aligned} E[D^{(S)}] &\leq \sum_{i=1}^M i P(i) + \sum_{i=M+1}^{\infty} \left( 1 + \left( \frac{i-(M+1)}{\beta} + 1 \right) \gamma \right) P(i) + R \\ &= \sum_{i=1}^M i P(i) + \sum_{i=M+1}^{\infty} \left( 1 \left( 1 + \frac{\gamma}{\beta} \right) + \left( 1 - \frac{M+1}{\beta} \right) \gamma \right) P(i) + R \\ &\leq \left( 1 + \frac{\gamma}{\beta} \right) E[\tilde{D}^{(S)}] + \left( 1 - \frac{M+1}{\beta} \right) \gamma + R \\ &= \frac{\alpha}{\beta} E[\tilde{D}^{(S)}] + ((\beta-1) - M) \left( \frac{\alpha}{\beta} - 1 \right) + R, \end{aligned}$$

which is the first part of the theorem.

By applying Relationship (5.47) of Theorem 5.4 to the second term of Equation (5.53), we have

$$\begin{aligned}
 E[D^{(S)}] &\geq \sum_{i=1}^M i P(i) + \sum_{i=M+1}^{\infty} \left(1 + \frac{i-M}{\beta} \gamma\right) P(i) + R \\
 &= \sum_{i=1}^{\infty} i P(i) + \sum_{i=M+1}^{\infty} \frac{i-M}{\beta} \gamma P(i) + R \\
 &\geq E[\tilde{D}^{(S)}] + \sum_{i=0}^{\infty} \frac{i-M}{\beta} \gamma P(i) + R \\
 &= E[\tilde{D}^{(S)}] + \frac{\gamma}{\beta} E[\tilde{D}^{(S)}] - \frac{M\gamma}{\beta} + R \\
 &= \frac{\alpha}{\beta} E[\tilde{D}^{(S)}] - M\left(\frac{\alpha}{\beta} - 1\right) + R,
 \end{aligned}$$

which is the second part of the theorem.

Q.E.D.

Note that

$$E[D^{(S)}]_u - E[D^{(S)}]_l = (\beta-1)\left(\frac{\alpha}{\beta} - 1\right), \quad (5.54)$$

and

$$E[D^{(S)}]_u = E[D^{(S)}]_l, \quad \text{if } \beta = 1. \quad (5.55)$$

#### Bounds on the Mean of D

We are now ready to obtain the mean of D by the following theorem.

#### Theorem 5.6

Under a Fixed Reservation scheme, if  $\rho < 1$ , the mean of D is upper bounded by

$$\begin{aligned}
E[D] \leq E[D^{(R)}] + \frac{\alpha}{\beta} & \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\
& \left. + \frac{1}{2} E[B] \left( \frac{E[A^2]}{E[A]} + 1 \right) \right\} \\
& + ((\beta-1) - M) \left( \frac{\alpha}{\beta} - 1 \right) + R,
\end{aligned} \tag{5.56}$$

and lower bounded by

$$\begin{aligned}
E[D] \geq E[D^{(R)}] + \frac{\alpha}{\beta} & \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\
& \left. - \frac{1}{2} E[G] + \frac{1}{2} E[B] \left( \frac{E[A^2]}{E[A]} + 1 \right) \right\} \\
& - M \left( \frac{\alpha}{\beta} - 1 \right) + R,
\end{aligned} \tag{5.57}$$

where

$$E[G] = E[A]E[B], \tag{5.58}$$

$$\rho = E[G]/\beta, \tag{5.59}$$

and

$$\text{Var}(G) = E[A]E[B^2] + E[B]^2(\text{Var}(A) - E[A]). \tag{5.60}$$

#### Proof

From Equation (5.1), we have

$$E[D] = E[D^{(R)}] + E[D^{(S)}]. \tag{5.61}$$

The proof is accomplished by applying the results of  $E[D^{(S)}]$ ,  $E[\tilde{D}^{(S)}]$  and  $E[W^{(L)}]$  successively into the above equation.

Q.E.D.

#### Fixed-Assignment TDMA Schemes

Recall that Fixed-Assignment TDMA schemes are actually pre-determined Fixed Reservation schemes. Hence, it is not necessary to

send reservation information. From Equation (5.4), we have

$$E[D^{(R)}] = (\alpha - 1)/2 . \quad (5.62)$$

Therefore the average message delay under a Fixed-Assignment TDMA scheme is the same as that described by Theorem 5.6 with  $E[D^{(R)}]$  substituted by Equation (5.62).

If  $\beta = 1$ , then  $M = 1$ . By applying Equation (5.37) into Theorem 5.6, we have

$$E[D] = \frac{\alpha}{2} \left\{ \frac{\text{Var}(G)}{1-\rho} + E[B] \frac{\text{Var}(A)}{E[A]} + E[B] \right\} - \frac{\alpha-1}{2} + R , \quad (5.63)$$

where

$$\rho = E[A]E[B] , \quad (5.64)$$

and

$$\text{Var}(G) = E[A]E[B^2] + E[B]^2(\text{Var}(A) - E[A]) . \quad (5.65)$$

Notice that Equation (5.63) is consistent with Equation (4.113) (see also [34]). Note that  $A$  in Equation (4.113) represents the number of messages arriving in a slot and  $A$  in Equation (5.63) represents the number of messages arriving in a frame,  $\alpha$  slots.

#### 5.4 Fixed Reservation Schemes with Preemptive Priorities

In the previous sections, we studied Fixed Reservation schemes under which all messages receive equal treatment. In this section, we introduce a preemptive priority discipline such that priority messages are treated before ordinary messages even if the former have successfully made reservations after the latter. The discipline is described as follows.



### Preemptive Resume Discipline

Messages arriving at sources are divided into two classes, Class 1 messages (priority messages), and Class 2 messages (ordinary messages). Class 1 messages which have successfully made reservations are transmitted before Class 2 messages, no matter when the Class 2 messages have successfully made their reservations. Upon receiving the reservation of a Class 1 message, if a Class 2 message is being transmitted, it is preempted for the transmission of the Class 1 message. A preempted Class 2 message resumes its transmission as soon as all the Class 1 messages which have successfully made reservations are completely transmitted.

Our analysis is based on the model under which all messages are divided into two classes. However, the results will be extended to the case where messages are divided into more than two classes by the following method. Suppose we want to investigate the delays of Class  $i$  messages, where smaller  $i$  for higher priority and larger  $i$  for lower priority. All messages of priorities higher than  $i$  can be regarded as Class 1 messages because messages of lower priorities are virtually invisible for them. Class  $i$  messages can be regarded as Class 2 messages. Other messages of lower priorities are virtually invisible for Class  $i$  messages. Hence, we only need to study the model under which there are two classes of messages. Before we investigate the message delays of both classes of messages, we describe the statistics of both classes of messages as follows.

The statistics of the two classes of messages are independent. We use indexes 1 and 2 to distinguish the two classes of messages.

### Assumption 5.3

The numbers of Class  $j$  messages,  $j = 1, 2$ , which have successfully made reservations in a sequence of reservation periods are independent and identically distributed. Let  $A_j$  be the number of Class  $j$  messages,  $j = 1, 2$ , which have successfully made reservations in a reservation period. The distribution of  $A_j$  is denoted by  $P_{A_j}(i)$  and the z-transform of  $P_{A_j}(i)$  is denoted by  $C_{A_j}(z)$ ,  $j = 1, 2$ .

### Assumption 5.4

The numbers of packets in Class  $j$  messages,  $j = 1, 2$ , are independent and identically distributed. Let  $B_j$  be the number of packets in a Class  $j$  message,  $j = 1, 2$ . The distribution of  $B_j$  is denoted by  $P_{B_j}(i)$  and the z-transform of  $P_{B_j}(i)$  is denoted by  $C_{B_j}(z)$ ,  $j = 1, 2$ .

### Class 1 Average Message Delay

In the order of transmission, Class 2 messages are invisible for Class 1 messages. Hence, the delay analysis for Class 1 messages is exactly the same as that for the messages in a system with only one class of messages. This has been studied in the previous sections and is not to be repeated here.

### Class 2 Average Message Delay

Before we proceed to study Class 2 message delays, we make the following definitions:

$G$  - the number of packets (both Class 1 and Class 2) which have successfully made reservations in a reservation period. Set  $P_G(k)$  to be the distribution of  $G$  and  $C_G(z)$  to be the z-transform of  $P_G(k)$ . From Assumptions 5.3 and 5.4, we have

$$P_G(k) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left( P_{B_1}^{(i)} * P_{B_2}^{(j)} \right) (k) P_{A_1}(i) P_{A_2}(j), \quad (5.66)$$

where  $\left( P_{B_1}^{(i)} * P_{B_2}^{(j)} \right) (k)$  is the  $i$ -time convolution of  $P_{B_1}(k)$  convolves with the  $j$ -time convolution of  $P_{B_2}(k)$ . Hence,

$$\begin{aligned} C_G(z) &= \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{B_1}^{(i)} * P_{B_2}^{(j)} (k) P_{A_1}(i) P_{A_2}(j) z^k \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_{B_1}(z)^i C_{B_2}(z)^j P_{A_1}(i) P_{A_2}(j) \\ &= C_{A_1} \left( C_{B_1}(z) \right) C_{A_2} \left( C_{B_2}(z) \right). \end{aligned} \quad (5.67)$$

$\rho$  - the throughput of the service portion of the channel, which is

$$\begin{aligned} \rho &= E[G]/\beta \\ &= (E[A_1]E[B_1] + E[A_2]E[B_2])/\beta. \end{aligned} \quad (5.68)$$

$V_n$  - the channel queue size immediately before the  $n^{\text{th}}$  reservation group is received completely. This is analogous to the term  $\tilde{W}_n^{(L)}$  defined in the previous sections.

The delay of a Class 2 message, denoted by  $D_2$ , can be split into two parts,

$$D_2 = D_2^{(R)} + D_2^{(S)}. \quad (5.69)$$

This is similar to Equation (5.1).  $D_2^{(R)}$  is the same as  $D^{(R)}$  in Equation (5.2) and is described by Equations (5.2) - (5.4). Similar to

Section 5.3,  $\tilde{D}_2^{(S)}$  is defined from  $D_2^{(S)}$  when only service slots are counted. Before we investigate  $\tilde{D}_2^{(S)}$ , we study  $V_n$ ,  $n > 0$ , as follows.

#### Bounds on the Mean of $V$

$V_n$  is the channel queue size immediately before the  $n^{\text{th}}$  reservation group is received completely. Set  $\tilde{G}_n$  to be the number of packets in the  $n^{\text{th}}$  group and  $\tau_n$  to be the number of reservation periods between the  $n^{\text{th}}$  and the  $n+1^{\text{th}}$  groups. We have

$$V_{n+1} = [V_n + \tilde{G}_n - \tau_n \beta]^+ . \quad (5.70)$$

#### Theorem 5.7

Under a Fixed Reservation scheme with preemptive priority, if  $\rho < 1$ , a limiting distribution of  $V_n$  exists. Its  $z$ -transform, denoted by  $C_V(z)$ , is given by

$$C_V(z) = \frac{\beta(1-\rho)(1-z)}{C_G(z) - z^\beta} \prod_{r=1}^{\beta-1} \frac{z - \eta_r}{1 - \eta_r} , \quad (5.71)$$

where  $\eta_r$ ,  $r = 1, 2, \dots, \beta-1$ , are the distinct  $\beta-1$  roots with  $|\eta_r| < 1$ , obtained by the limit

$$\eta_r = \lim_{\omega \rightarrow 1^-} \eta_r(\omega) , \quad (5.72)$$

where  $\eta_r(\omega)$ ,  $r = 1, 2, \dots, \beta$ , are the  $\beta$  distinct roots of the functional equation

$$z^\beta = \omega C_G(z) . \quad (5.73)$$

Furthermore,

$$E[V] = \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] + \frac{1}{2} \sum_{r=1}^{\beta-1} \frac{1 + \eta_r}{1 - \eta_r} \quad (5.74)$$

Proof

Similar technique as Theorem 5.1.

Q.E.D.

Theorem 5.8

If  $\rho < 1$ , the Markov Chain  $\{V_n, n > 0\}$  has a steady state mean,  $E[V]$ , which is upper bounded by

$$E[V]_u \triangleq \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \quad (5.75)$$

and lower bounded by

$$E[V]_l \triangleq \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} - \frac{1}{2} E[G] \quad (5.76)$$

Proof

Similar technique as Theorem 5.2.

Q.E.D.

Note that

$$E[V]_u - E[V]_l = \frac{1}{2} E[G] \quad (5.77)$$

Also note that

$$E[V] = E[V]_l, \quad \text{if } \beta = 1. \quad (5.78)$$

Bounds on the mean of  $\tilde{D}_2^{(S)}$

We are now ready to investigate  $\tilde{D}_2^{(S)}$  as follows.

Theorem 5.9

Under a Fixed Reservation scheme with preemptive priority, if  $\rho < 1$ , the mean of  $\tilde{D}_2^{(S)}$  is upper bounded by

$$\begin{aligned} E[\tilde{D}_2^{(S)}]_u \triangleq & \left\{ E[V] + E[A_1]E[B_1] \left(1 - \frac{1}{\beta}\right) \right. \\ & \left. + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \right\} \\ & \cdot (1 - E[A_1]E[B_1]/\beta)^{-1}, \end{aligned} \quad (5.79)$$

and lower bounded by

$$E[\tilde{D}_2^{(S)}] \stackrel{\Delta}{=} \left\{ E[V] + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \right\} \cdot (1 - E[A_1]E[B_1]/\beta)^{-1} . \quad (5.80)$$

Proof

$\tilde{D}_2^{(S)}$  can be decomposed into

$$\tilde{D}_2^{(S)} = B_2 + V + G_1 + \tilde{W}_2^{(G)} + F(\tilde{D}_2^{(S)}) . \quad (5.81)$$

The first term,  $B_2$ , is the service time of a Class 2 message. The second term,  $V$ , is the channel queue size  $V_n$ ,  $n > 0$ , at steady state. The third term,  $G_1$ , is the Class 1 messages which have arrived in the same reservation period as did the Class 2 message under consideration. The fourth term,  $\tilde{W}_2^{(G)}$ , is the number of Class 2 packets which are in the same group, but transmitted before the Class 2 message under consideration. The last term,  $F(\tilde{D}_2^{(S)})$ , is the number of Class 1 packets which arrive during  $\tilde{D}_2^{(S)}$ .

Consider the last term. If  $j\beta < \tilde{D}_2^{(S)} \leq (j+1)\beta$ ,  $j \geq 0$ , then the number of Class 1 messages, denoted by  $U_j$ , which arrive during  $\tilde{D}_2^{(S)}$  is

$$U_j = \sum_{k=1}^j A_{1k} , \quad (5.82)$$

where  $A_{1k}$  is the number of Class 1 messages which arrive in the  $k^{\text{th}}$  reservation period during  $\tilde{D}_2^{(S)}$ . Hence, we have

$$F(\tilde{D}_2^{(S)}) = \sum_{i=1}^{U_j} B_{1i} , \quad (5.83)$$

where  $B_{1i}$  is the number of packets in the  $i^{\text{th}}$  Class 1 message which

arrives during  $\tilde{D}_2^{(S)}$ . Set  $P(i)$  to be the distribution of  $\tilde{D}_2^{(S)}$ . By taking expecting values on both sides of Equation (5.83), we have

$$\begin{aligned} E[F(\tilde{D}_2^{(S)})] &= \sum_{j=0}^{\infty} j E[A_1] E[B_1] \sum_{i=j\beta+1}^{(j+1)\beta} P(i) \\ &= E[A_1] E[B_1] \sum_{j=0}^{\infty} \sum_{i=j\beta+1}^{(j+1)\beta} j P(i) . \end{aligned} \quad (5.84)$$

By applying Relationship (5.46) of Theorem 5.4 to the right side of Equation (5.84), we have

$$\begin{aligned} E[F(\tilde{D}_2^{(S)})] &\leq E[A_1] E[B_1] \sum_{i=1}^{\infty} \frac{i-1}{\beta} P(i) \\ &= E[A_1] E[B_1] (E[\tilde{D}_2^{(S)}] - 1) / \beta . \end{aligned} \quad (5.85)$$

By applying Relationship (5.47) of Theorem 5.4 to the right side of Equation (5.84), we have

$$\begin{aligned} E[F(\tilde{D}_2^{(S)})] &\geq E[A_1] E[B_1] \sum_{i=1}^{\infty} \left( \frac{1}{\beta} - 1 \right) P(i) \\ &= E[A_1] E[B_1] \left( \frac{E[\tilde{D}_2^{(S)}]}{\beta} - 1 \right) . \end{aligned} \quad (5.86)$$

Taking expectations on both sides of Equation (5.81), we have

$$\begin{aligned} E[\tilde{D}_2^{(S)}] &= E[B_2] + E[V] + E[G_1] + E[W_2^{(G)}] \\ &\quad + E[F(\tilde{D}_2^{(S)})] . \end{aligned} \quad (5.87)$$

From Assumptions 5.3 and 5.4, the third term is

$$E[G_1] = E[A_1] E[B_1] \quad (5.88)$$

By applying Theorem 5.3 to the fourth term, we have

$$E[\tilde{W}_2^{(G)}] = \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} - 1 \right). \quad (5.89)$$

Hence,

$$\begin{aligned} E[\tilde{D}_2^{(S)}] &= E[B_2] + E[V] + E[A_1]E[B_1] \\ &\quad + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} - 1 \right) + E[F(\tilde{D}_2^{(S)})]. \end{aligned} \quad (5.90)$$

By applying Relationship (5.85) into the above equation, we have

$$\begin{aligned} E[\tilde{D}_2^{(S)}] &\leq E[V] + E[A_1]E[B_1] + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \\ &\quad + E[A_1]E[B_1](E[\tilde{D}_2^{(S)}] - 1)/\beta. \end{aligned} \quad (5.91)$$

Rearrange terms, we have

$$\begin{aligned} E[\tilde{D}_2^{(S)}] &\leq E[\tilde{D}_2^{(S)}]_u = \left\{ E[V] + E[A_1]E[B_1](1 - 1/\beta) \right. \\ &\quad \left. + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \right\} (1 - E[A_1]E[B_1]/\beta)^{-1}. \end{aligned}$$

By applying Relationship (5.86) into Equation (5.90), we have

$$\begin{aligned} E[\tilde{D}_2^{(S)}] &\geq E[V] + E[A_1]E[B_1] + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \\ &\quad + E[A_1]E[B_1] \left( \frac{E[\tilde{D}_2^{(S)}]}{\beta} - 1 \right). \end{aligned} \quad (5.92)$$

Rearrange terms, we have



$$E[\tilde{D}_2^{(S)}] \geq E[\tilde{D}_2^{(S)}]_l = \left\{ E[V] + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \right\} \cdot (1 - E[A_1]E[B_1]/\beta)^{-1}.$$

Q.E.D.

Note that

$$E[\tilde{D}_2^{(S)}]_u - E[\tilde{D}_2^{(S)}]_l = E[A_1]E[B_1](1 - 1/\beta) \cdot (1 - E[A_1]E[B_1]/\beta)^{-1}, \quad (5.93)$$

and

$$E[\tilde{D}_2^{(S)}]_u = E[\tilde{D}_2^{(S)}]_l, \text{ if } \beta = 1. \quad (5.94)$$

#### Bounds on the Mean of $D_2$

We are now ready to state the mean of  $D_2$  as follows.

#### Theorem 5.10

Under a Fixed Reservation scheme, if  $\rho < 1$ , the mean of  $D_2$  is upper bounded by

$$\begin{aligned} E[D_2] \leq E[D^{(R)}] + \frac{\alpha}{\beta(1-E[A_1]E[B_1]/\beta)} & \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\ & + E[A_1]E[B_1](1 - 1/\beta) + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \left. \right\} \\ & + ((\beta-1) - M) \left( \frac{\alpha}{\beta} - 1 \right) + R, \end{aligned} \quad (5.95)$$

and lower bounded by

$$\begin{aligned}
E[D_2] &\geq E[D^{(R)}] + \frac{\alpha}{\beta(1-E[A_1]E[B_1]/\beta)} \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\
&\quad \left. - \frac{1}{2} E[G] + \frac{1}{2} E[B_2] \left( \frac{E[A_2^2]}{E[A_2]} + 1 \right) \right\} \\
&\quad - M \left( \frac{\alpha}{\beta} - 1 \right) + R,
\end{aligned} \tag{5.96}$$

where

$$E[G] = \sum_{j=1}^2 E[A_j]E[B_j], \tag{5.97}$$

$$\rho = E[G]/\beta, \tag{5.98}$$

and

$$\begin{aligned}
\text{Var}(G) &= \sum_{j=1}^2 \left\{ E[A_j]E[B_j^2] \right. \\
&\quad \left. + E[B_j]^2(\text{Var}(A_j) - E[A_j]) \right\}.
\end{aligned} \tag{5.99}$$

#### Proof

From Equation (5.69), we have

$$E[D_2] = E[D^{(R)}] + E[D_2^{(S)}]. \tag{5.100}$$

By applying the results of  $E[V]$ ,  $E[\tilde{D}^{(S)}]$  is obtained which is then converted to  $E[D^{(S)}]$  by Theorem 5.5 and the proof is completed.

Q.E.D.

#### Class i Average Message Delay

So far, we have studied the case in which there are two classes of messages. We now extend our results to the case in which there are more than two classes of messages.

Theorem 5.11

Under a Fixed Reservation scheme with preemptive priorities, if  $\rho < 1$ , the average Class 1 message delay, denoted by  $E[D_1]$ ,  $1 > 0$ , is upper bounded by

$$\begin{aligned} E[D_1] \leq E[D^{(R)}] + \frac{\alpha}{\beta(1-E[G_P]/\beta)} \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\ \left. + E[G_P](1-1/\beta) + \frac{1}{2} E[B_1] \left( \frac{E[A_1^2]}{E[A_1]} + 1 \right) \right\} \\ + ((\beta-1)-M) \left( \frac{\alpha}{\beta} - 1 \right) + R, \end{aligned} \quad (5.101)$$

and lower bounded by

$$\begin{aligned} E[D_1] \geq E[D^{(R)}] + \frac{\alpha}{\beta(1-E[G_P]/\beta)} \left\{ \frac{1}{2} \frac{\text{Var}(G)}{\beta(1-\rho)} \right. \\ \left. - \frac{1}{2} E[G] + \frac{1}{2} E[B_1] \left( \frac{E[A_1^2]}{E[A_1]} + 1 \right) \right\} \\ - M \left( \frac{\alpha}{\beta} - 1 \right) + R, \end{aligned} \quad (5.102)$$

where

$$E[G_P] = \sum_{j=1}^{1-1} E[A_j]E[B_j], \quad (5.103)$$

$$E[G] = E[G_P] + E[A_1]E[B_1], \quad (5.104)$$

$$\rho = E[G]/\beta, \quad (5.105)$$

and

$$\begin{aligned} \text{Var}(G) = \sum_{j=1}^1 \left\{ E[A_j]E[B_j^2] \right. \\ \left. + E[B_j]^2 (\text{Var}(A_j) - E[A_j]) \right\}. \end{aligned} \quad (5.106)$$

### Proof

The proof is accomplished by applying induction to Theorem 5.10.

Q.E.D.

Note that Expressions (5.101) - (5.106) are identical to Expressions (5.56) - (5.60), if  $i=1$ .

### Fixed-Assignment TDMA Schemes

Under Fixed-Assignment TDMA schemes, it is not necessary for messages to send reservation information. Hence,

$$E[D^{(R)}] = \frac{\alpha-1}{2} . \quad (5.107)$$

$E[D_i]$  is then given by Theorem 5.11 with the term  $E[D^{(R)}]$  substituted by the above value.

If  $\beta = 1$ , then  $M = 1$ . By applying Equation (5.78) into Theorem 5.11, Relationships (5.101) - (5.106) become

$$\begin{aligned} E[D_1] = & \frac{\alpha}{1-E[G_p]} \left\{ \frac{1}{2} \frac{\text{Var}(G)}{(1-\rho)} \right. \\ & - \frac{1}{2} E[G] + \frac{1}{2} E[B_1] \left( \frac{E[A_1^2]}{E[A_1]} + 1 \right) \Big\} \\ & - \frac{\alpha-1}{2} + R, \end{aligned} \quad (5.108)$$

where

$$E[G_p] = \sum_{j=1}^{i-1} E[A_j] E[B_j] , \quad (5.109)$$

$$E[G] = E[G_p] + E[A_1] E[B_1] \quad (5.110)$$

$$\rho = E[G]/\beta , \quad (5.111)$$

and

$$\begin{aligned} \text{Var}(G) = & \sum_{j=1}^1 \left\{ E[A_j] E[B_j^2] \right. \\ & \left. + E[B_j]^2 (\text{Var}(A_j) - E[A_j]) \right\} . \end{aligned} \quad (5.112)$$

### Numerical Results

To demonstrate the effect of employing a preemptive priority discipline under a Fixed Reservation scheme, we consider a Fixed-Assignment TDMA scheme with  $\alpha = 10$  (slots),  $\beta = 1$  (slot) and  $R = 12$  (slots). Class 1 and Class 2 messages arrive independently according to Poisson distributions with parameters  $\lambda_1$  (messages per frame) and  $\lambda_2$  (messages per frame), respectively. The lengths of both classes of messages are statistically independent and both follow Geometric distributions with means  $1/p_1$  (packets per message), and  $1/p_2$  (packets per message), respectively. Hence,

$$E[G_p] = \lambda_1/p_1,$$

$$E[G] = \sum_{j=1}^2 \lambda_j/p_j ,$$

$$\rho = \sum_{j=1}^2 \lambda_j/p_j ,$$

and

$$\text{Var}(G) = \sum_{j=1}^2 \lambda_j (2-p_j)/p_j^2$$

If  $\lambda_1 = 0.3$  and  $p_1 = 1.0$ , we have  $E[D_1] = 19.64$  (slots). If  $\lambda_1 = 0.5$  and  $p_1 = 1.0$ , we have  $E[D_1] = 22.5$  (slots). Set  $p_2 = 0.25$  and  $\rho_j = \lambda_j/p_j$ ,  $j = 1, 2$ . In Figure 5.1,  $E[D_2]$  is plotted versus  $\rho_2$  when

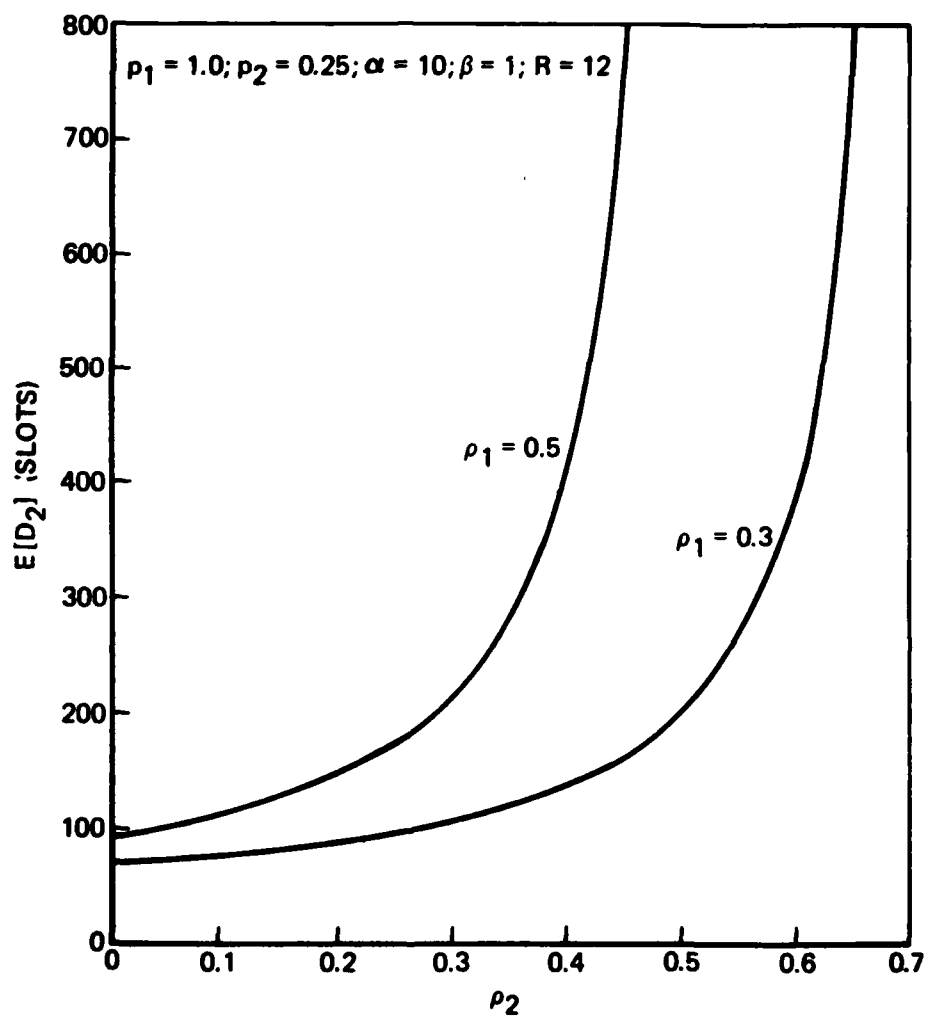


Figure 5.2. Class 2 Average Message Delay vs  $\rho_2$  under a Preemptive Reservation Scheme.

$\rho_1 = 0.3$  and when  $\rho_1 = 0.5$ , respectively. These curves clearly illustrate the effect of Class 1 message traffic on Class 2 average message delay.

## CHAPTER VI

### CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

#### 6.1 Conclusions

A TDMA (Time-Division Multiple-Access) scheme is used to control the sharing of a communication channel on a time-division basis by geographically distributed sources. To completely describe and analyze a TDMA system, we need to specify the communication channel, the characteristics of the sources that share the channel, the statistics of the messages arriving at the sources and the access-control scheme that allocates the channel capacity among the sources. Several TDMA systems have been studied in this dissertation.

In systems where sources require real-time transmission at specific information rates, Demand-Assignment TDMA schemes are employed to allocate available slots to sources that have messages to transmit. Two cases have been studied.

In one case, all sources are assumed to require transmission at one specific information rate. A Cutoff Priority discipline has also been introduced to give advantage to more important messages. A numerical example has been presented and performance curves have been plotted and compared.

In the other case, sources are assumed to require real-time transmission at multiple information rates. An overall system performance measure is proposed. Based on this measure, the capacity of the channel is optimally allocated among the sources. A numerical example has been presented and the procedure of finding the optimal



frame structure has been demonstrated.

In systems where sources do not require real-time transmission at specific information rates, a store-and-forward technique is employed and messages are transmitted at full channel capacity when they reach the head of the channel queue. We have studied a class of hybrid TDMA/Collision-Resolving scheme and a class of Fixed Reservation schemes in allocating channel capacity, on a store-and-forward basis, among the sources that have messages to transmit.

The class of hybrid TDMA/Collision-Resolving schemes is composed of a Fixed-Assignment TDMA component and random access collision-resolving components. In systems where messages are short single packets, this class of schemes have been shown to yield good delay-throughput characteristics. Two families of this class of schemes, developed from two versions of a Tree Search random access scheme, have been analyzed. These two families of schemes have been applied separately to a system in which the sources are assumed to have limited buffer capacities. Numerical examples have been given and performance curves have been plotted and compared. These two families of schemes have then been applied to another system in which the sources are assumed to have unlimited buffer capacities. Again, numerical examples have been given and performance curves have been plotted and compared.

Finally, a class of Fixed Reservation access-control schemes have been studied. A mathematical technique (Theorem 5.4) has been developed to yield upper and lower bounds on the average message delay. A preemptive priority discipline has been incorporated as an important and practical feature in the operation of the Fixed Reservation scheme

analyzed here. Bounds on the average delays of messages of different priorities have been derived. In some special cases, these bounds converge and yield exact results. This has been demonstrated by numerical examples.

## 6.2 Suggestions for Future Research

Under each of the access-control scheme we have studied, the sources are considered to have the same statistical message characteristics. In realistic situations, the sources that share a communication channel are not all alike and the messages arriving at the sources are of different characteristics. Messages arriving at real-time sources, such as vocoders and teleprinters, may require real-time transmission at different information rates. Interactive messages, such as user to computer and user to user messages, may require short response times at low or high information rates. Batch data messages, such as in document distribution, may have longer prescribed response times but can be subject to preemption. Hence, one should incorporate the appropriate combinations of access-control schemes and priority disciplines when designing and developing communication networks to meet a spectrum of performance requirements in accommodating messages of diverse characteristics.

APPENDIX  
THE APPROXIMATION OF A DEMAND-ASSIGNMENT TDMA SYSTEM  
BY AN M/M/K QUEUEING SYSTEM

The difference between an M/M/K queueing system and a Demand-Assignment TDMA system can be described as follows.

In an M/M/K queueing system, all the customers who are granted servers are served simultaneously in time. Also, a customer who is granted a server will start his service immediately.

Under a Demand-Assignment TDMA scheme, all the sources that are granted channel slots transmit simultaneously, so that each of them is dedicated one single slot in every frame. However, at any given time, only one single source can transmit. Also, a message which is granted a slot usually cannot transmit immediately. It suffers a frame latency delay, which is measured from the time of assignment of a slot to the time transmission actually starts. (This is illustrated by Figure A.)

Assume that under the Demand-Assignment TDMA scheme, the overall transmission times of messages over the channel are statistically independent and identically distributed, with an exponential distribution. Let  $B$  (seconds) be the transmission time of a message with mean  $\mu_M^{-1}$  (seconds). Denote the distribution of  $B$  by  $P_B(t)$ . Then

$$P_B(t) = \mu_M^{-1} e^{-\mu_M t}, \quad t \geq 0. \quad (A.1)$$

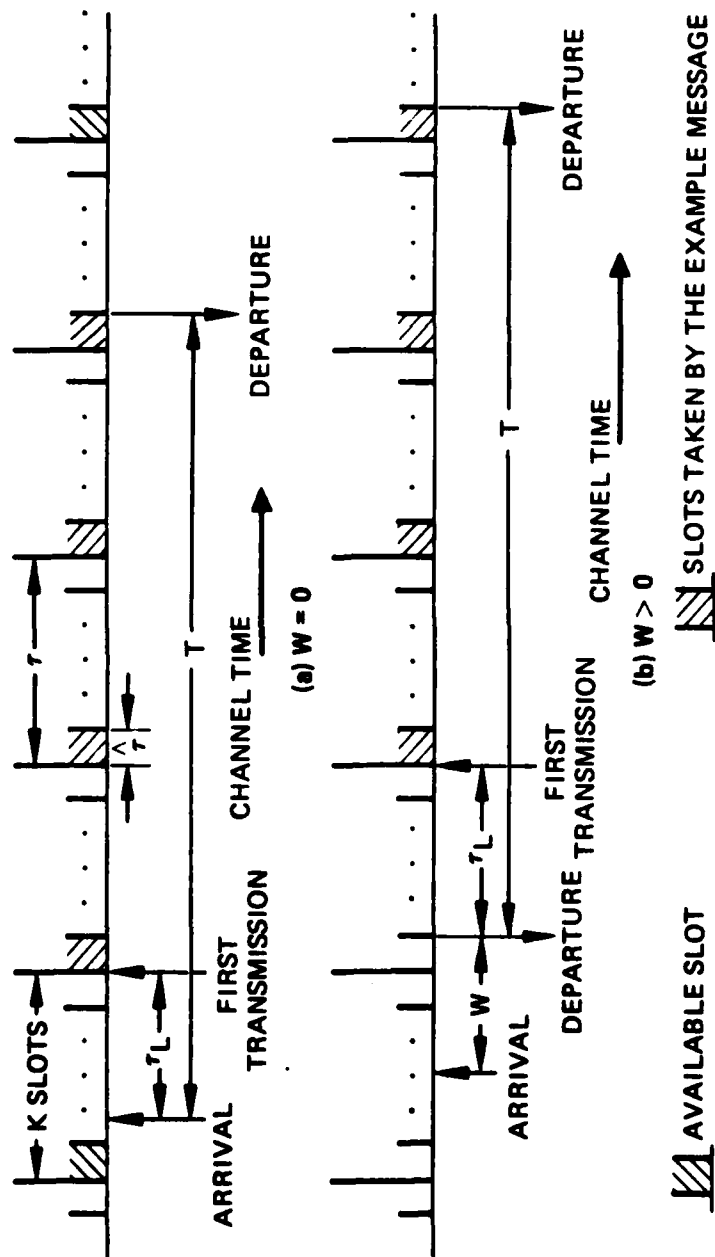


Figure A. The Frame Latency,  $\tau_L$ , and the Holding Time,  $T$ , of a Message under a DA/TDMA Scheme, for (a)  $W = 0$  and (b)  $W > 0$ .

Under the Demand-Assignment TDMA scheme, define the holding time,  $T$  (seconds), of a message to be measured from the time a slot is assigned to the message until the time representing the end of the last slot used by the message. Also let  $W$  be the waiting time of the message (from the time of its arrival until the time a slot is assigned to it). We can express  $T$  in terms of  $B$  as

$$T = \begin{cases} \tau_L + [B/\tau]\tau + \tau & \text{if } B/\tau \text{ is not an integer,} \\ \tau_L + [B/\tau]\tau - (\tau - \tau) & \text{if } B/\tau \text{ is an integer,} \end{cases} \quad (\text{A.2})$$

where  $[X]$  is the integer of  $X$ .

From Equation (A.2), we conclude that

$$\tau_L + KB - (\tau - \tau) \leq T < \tau_L + KB + \tau \quad (\text{A.3})$$

If  $W = 0$ ,  $0 \leq \tau_L \leq \tau$  and Relationship (A.3) becomes

$$KB - (\tau - \tau) \leq T < KB + \tau + \tau \quad (\text{A.4})$$

If  $W > 0$ , then  $\tau_L = \tau - \tau$  and Relationship (A.3) becomes

$$KB \leq T < KB + \tau \quad (\text{A.5})$$

From Relationships (A.4) and (A.5), we have

$$KB - (\tau - \tau) \leq T < KB + \tau + \tau \quad (\text{A.6})$$

Hence, if  $KB \gg \tau$ ,  $T \approx KB$ .

Consequently, if  $KE[B] \gg \tau$ , we conclude that an M/M/K queueing system well approximates a Demand-Assignment TDMA system. This M/M/K queueing system has an arrival process identical to that of the

Demand-Assignment TDMA system and service time distribution, denoted by  $P_T(t)$ , given by

$$P_T(t) = \mu e^{-\mu t}, \quad t \geq 0, \quad (A.7)$$

where

$$\mu = \mu_M/K. \quad (A.8)$$

## BIBLIOGRAPHY

1. Huff, R., "TDMA Space Communications Systems: Concepts and Practical Techniques," Rome Air Development Central Technical Report, RADC-TR-71-255, November 1971.
2. Frenkel, G., G. Kinal, D. Gan and J. Doggett, "TDMA Network Control Study," Rome Air Development Center Technical Report, RADC-TR-73-270, August 1973.
3. Doggett, J., G. Kinal and S. Talbot, "Control of Channel Assignment in Tactical TDMA Satellite Communications," Proceedings of the National Telecommunication Conference, Vol. 1, 1973.
4. Rubin, I., "Message Delays in FDMA and TDMA Communication Channels," UCLA Technical Report, UCLA-ENG-7826, February 1978. Also, Proceedings of the Conference on Information Sciences and Systems, John Hopkins University, Baltimore, MD, March 1978.
5. Louie, M. and I. Rubin, "On Access Control disciplines for a TDMA System with Multiple-Rate Real-Time Sources," Proceedings of the International Telemetry Conference, Los Angeles, September 1976, Vol. XII, pp. 326-336.
6. Kleinrock, L., Communication Nets: Stochastic Message Flow and Delay, McGraw Hill, New York, 1964, out of print (Reprinted by Dover Publications, New York, 1972).
7. Abramson, N., "THE ALOHA SYSTEM - Another Alternative for Computer Communications," Proceedings of the AFIPS Fall Joint Computer Conference, 1970, Vol. 37, pp. 281-285.
8. Roberts, L.G., "ALOHA Packet System With and Without Slots and Capture," ARPA Network Information Center, Stanford Research Institute, Menlo Park, California, ASS Note 8 (NIC 11290), June 1972.
9. Capetanakis, J.I., The Multiple Access Broadcast Channel: Protocol and Capacity Considerations, Ph.D. Thesis, MIT, 1977.
10. Rubin, I., "Group Random Access Disciplines for Multi-Access Broadcast Channels," IEEE Transactions on Information Theory, Vol. IT-24, No. 5, pp. 578-592, September 1978.
11. Abramson, N., "Packet Switching with Satellites," Proceedings of the AFIPS National Computer Conference, New York, June 1973, Vol. 42, pp. 695-702.

12. Metcalfe, R.M., "Steady-State Analysis of a Slotted and Controlled ALOHA System with Blocking," Proceedings of the Sixth Hawaii International Conference on System Sciences, Honolulu, January 1973, pp. 375-387.
13. Metcalfe, R.M., "Packet Communication," MIT Technical Report, MAC TR-114, July 1973.
14. Kleinrock, L. and S.S. Lam, "Packet Switching in a Multi-Access Broadcast Channel: Performance Evaluation," IEEE Transactions on Communications, Vol. 23, No. 4, pp. 410-423, April 1975.
15. Lam, S.S., "Packet Switching in a Multiple-Access Broadcast Channel: Dynamic Control Procedures," IEEE Transactions on Communications, Vol. 23, No. 9, pp. 891-904, September 1975.
16. Kleinrock, L. and F. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple Access Models and Their Throughput Delay Characteristics," IEEE Transactions on Communications, Vol. COM-23, pp. 1400-1416, December 1975.
17. Tobagi, F. and L. Kleinrock, "Packet Switching in Radio Channels: Part II - The Hidden Terminal Problem in Carrier Sense Multiple Access and the Busy Tone Solution," IEEE Transactions on Communications, Vol. COM-23, pp. 1417-1433, December 1975.
18. Tobagi, F. and L. Kleinrock, "Packet Switching in Radio Channels: Part III - Polling and (Dynamic) Splitting-Channel Reservation Multiple Access," IEEE Transactions on Communications, Vol. COM-24, pp. 832-844, August 1976.
19. Tobagi, F. and L. Kleinrock, "Packet Switching in Radio Channels: Part IV - Stability Considerations and Dynamic Control in Carrier Sense Multiple Access," IEEE Transactions on Communications, Vol. COM-25, No. 10, pp. 1103-1119, October 1977.
20. Kleinrock, L. and Y. Yemini, "An Optimal Adaptive Scheme for Multiple Access Broadcast Communication," Conference Record International Conference on Communications, Toronto, June 1978, pp. 7.2.1-7.2.5.
21. Crowther, W., R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A System for Broadcast Communication: Reservation ALOHA," Proceedings of the Sixth International Conference on System Sciences, Honolulu, January 1973, pp. 371-374.
22. Roberts, L.G., "Dynamic Allocation of Satellite Capacity through Packet Reservation," Proceedings of the AFIPS National Computer Conference, New York, June 1973, Vol. 42, pp. 711-716.



23. Binder, R., "A Dynamic Packet-Switching System for Satellite Broadcast Channels," University of Hawaii Technical Report, B74-5, August 1974.
24. Rubin, I., "Access Control Disciplines for Multiple-Access Broadcast Channels: Reservation and TDMA Schemes," UCLA Technical Report, UCLA-ENG-7825, February 1978. Also, IEEE Transactions on Information Theory, Vol. IT-25, No. 5, pp. 516-536, September 1979.
25. Rubin, I., "Integrated Random Access Reservation Schemes for Multi-Access Communication Channels," UCLA Technical Report, UCLA-ENG-7752, July 1977.
26. Kleinrock, L. Queueing Systems, Vol. 1: Theory, John Wiley and Sons, New York, 1975.
27. Jaiswal, N., Priority Queues, Academic Press, New York, 1968.
28. Chung, K.L., Markov Chains with Stationary Transition Probabilities, Springer-Verlag, New York, 1967.
29. Leibowitz, M., "An Approximation Method for Treating a Class of Multiqueue Problems," IBM Journal, Vol. 5, No. 3, pp. 204-209, July 1961.
30. Hashida, O., "Analysis of Multiqueue," Nippon Telegraph and Telephone Public Corporation, Review of the Electric Communication Laboratories, Vol. 20, No. 3-4, pp. 189-199, March-April 1972.
31. Hashida, O. and K. Ohara, "Line Accommodation Capacity of a Communication Control Unit," Nippon Telegraph and Telephone Public Corporation, Review of the Electric Communication Laboratories, Vol. 20, No. 3-4, pp. 231-239, March-April 1972.
32. Kingman, J., "Some Inequalities for the Queue GI/G/1," Biometrika, 49, 1962, pp. 315-324.
33. Cohen, J., The Single Server Queue, North-Holland Publishing Company, Amsterdam, 1969.
34. Lam, S.S., "Delay Analysis of a Time Division Multiple Access (TDMA) Channel," IEEE Transactions on Communications, Vol. COM-25, pp. 1489-1494, December 1977.